

## When to believe the unbelievable

*An article in this week's issue describes observations for which there is no present physical basis. There are good and particular reasons why prudent people should, for the time being, suspend judgement.*

INEXPLICABLE observations are not always signs of the supernatural. That is what readers of the remarkable article on page 816 should keep in mind. They should also remember that Avogadro's number, the number of molecules in a gram molecule of material, is roughly  $6.23 \times 10^{23}$ , which naturally implies that most of the experiments with antibody solution reported by Professor J. Benveniste and his colleagues have been carried out in the literal absence of antibody molecules. For what the article shows is that it is possible to dilute an aqueous solution of an antibody virtually indefinitely without the solution losing its biological activity. Or rather, there is a surprising rhythmic fluctuation in the activity of the solution. At some dilutions, the activity falls off; on further dilution, it is restored.

There is no objective explanation of these observations. Nor is there much comfort for anybody in the explanation offered at the end of the article — that antibody molecules once embodied in water leave their internal marks, as ghosts of a kind, on its molecular structure — for there is no evidence of any other kind to suggest that such behaviour may be within the bounds of possibility. Indeed, during the long period since this article was first submitted to *Nature*, it has been plain that Benveniste has been as puzzled as many of those who have read his article by the data he reports. On many occasions, he has responded to referees' suggestions at great inconvenience to himself. When told, for example, that the experiments should be repeated at an independent laboratory, he arranged for this to be done.

One of the purposes that will be served by publishing the article will be to provide an authentic account of this work for the benefit of those, especially in France, who have gathered rumours of it from the popular press. Another is that vigilant members of the scientific community with a flair for picking holes in other people's work may be able to suggest further tests of the validity of the conclusions.

Certainly there can be no justification, at this stage, for an attempt to use Benveniste's conclusions for the malign purposes to which they might be put. There are some obvious dangers. In homoeopathic medicine, for example, which works on the principle that very small concentrations of appropriate products may have consequences that far outweigh those expected of them, there will be a natural inclination to welcome Benveniste's article as aid and comfort, but that would be premature, probably mistaken. It will be time for celebrations of that kind only when a lot more water has run underneath this bridge.

But, those of supernatural inclinations will protest, is it not grossly unfair that science should put aside, even temporarily, some surprising and unexpected observations (such as these) while apparently welcoming others which are no less surprising (such as the recent suggestion that there may be a 'fifth force' between material objects)? The explanation is simple, but, perhaps for that reason, not widely understood. It is entirely possible for physicists to welcome the notion of the fifth force because it would be a novel happening which could nevertheless be accommodated within the accepted framework of science. Benveniste's observations, on the other hand, are startling not merely because they point to a novel phenomenon, but because they strike at the roots of two centuries of observation and

rationalization of physical phenomena. Where, for example, would elementary principles such as the Law of Mass Action be if Benveniste is proved correct? The principle of restraint which applies is simply that, when an unexpected observation requires that a substantial part of our intellectual heritage should be thrown away, it is prudent to ask more carefully than usual whether the observation may be incorrect. □

## Criminalizing research

*West Germany seems bent on a restrictive law on embryo research.*

THE efforts of West Germany's scientific organizations — Max Planck Society, Deutsche Forschungsgemeinschaft (DFG) — to prevent the passage of a restrictive Embryo Protection Law have apparently missed the mark (see page 791). Even if there were a party brave enough to amend the Justice Ministry's proposed ban on embryo research, it is hard to imagine that it could win a majority. More probably, given the prevailing mood in West Germany that "One shouldn't meddle with the Lord's handiwork", such a course would invite electoral trouble. Yet West Germany appears content with legal abortion: an eight-hour debate at the ruling Christian Democrats' party congress in mid-June led merely to an affirmation of the *status quo*. Is it not a paradox that research should be singled out as a scapegoat?

Much the same, of course, is happening in Britain, where the promised bill on embryo research seems certain to feature in the next session of Parliament, and where abortion remains a contentious issue. But there are two reasons why West Germany may be impelled towards meticulous central regulation rather than allowing researchers to police themselves.

First, there are the abuses of the past. Contemporary researchers may complain that they are being punished for acts for which they are not responsible, but, sadly, West German science, like its foreign relations and its military policy, still lives in the shadow of Nazi crimes. Second, the West German legal system differs dramatically from the Anglo-Saxon system. Whereas the Anglo-Saxon system applies a "principle of opportunity" allowing prosecutors broad discretion in decisions whether or not to prosecute a case, the German system forces a prosecutor to act if there is even a suspicion that a law has been broken. West German courts are not bound by legal precedents, as in Britain or the United States, so that legislators have no choice but to express their intentions in the minutest detail.

It will be small consolation for West German researchers that their work may be curtailed because of history and the legal system. At first, the Max Planck society and the DFG seemed incredulous that such a thing could happen; after all, "freedom of research" was guaranteed by the West German constitution. But now these organizations realize that research takes second place to human dignity, whatever that may mean. That, at least, is how the Justice Ministry sees the issue. But how soon will it be before people in West Germany appreciate that embryo research may, in due course, enhance that same ideal, perhaps by helping to rid people of undignifying genetic diseases? □



# NIH speeds to resolve charges of scientific error

- Congressional inquiry spurs action
- Nobel prizewinner defends research

## Washington

THE National Institutes of Health (NIH) has moved into high gear in its investigation of a paper by researchers at the Whitehead Institute of the Massachusetts Institute of Technology (MIT) and Tufts University who include the Nobel laureate David Baltimore.

A three-member team of immunologists appointed by NIH was in Boston last week conducting interviews. NIH director James Wyngaarten expects their report will be on his desk within weeks.

The dispute over the paper (*Cell* 45, 247; 1986) has become a focus of media attention since allegations of data misrepresentation in the paper, made by two NIH researchers, Walter Stewart and Ned Feder, were taken up by a Congressional subcommittee hearing on scientific fraud and misconduct (see *Nature* 332, 671, 1988).

Stewart and Feder are well known for their controversial investigations into scientific misconduct. Their criticisms of the *Cell* paper have been widely circulated and an analysis of them appears on pages 795–797 of this issue.

David Baltimore responded with a "Dear Colleague" letter to the scientific community, dated 17 May, that defends the research in question and urges a speedy review by NIH. Baltimore says he has consistently sought an investigation by researchers with sufficient expertise to evaluate the issues, something he believes Stewart and Feder do not possess.

Congressional interest in the veracity of a paper published in a specialist journal has many in the scientific community worried. The committee involved — the oversight and investigations subcommittee of the House of Representatives Committee on Energy and Commerce — has a powerful chairman in John Dingell (Democrat, Michigan) and has developed a reputation for tough questioning and 'hardball' tactics. But it has more expertise in investigating defence contracts than in checking data on the expression of immunoglobulin genes.

In its hearing on scientific misconduct it concentrated on the *Cell* paper. But Baltimore was not invited to testify, nor were any of the other five authors, including the principal author, Thereza Imanishi-Kari, now at Tufts University.

The chief witnesses were Stewart, Feder and O'Toole, all on record as critics of the paper. The subcommittee also did

not contact members of a Tufts University committee that had already investigated the paper.

For his part, Baltimore writes that "The halls of Congress are not the place to determine scientific truth or falsity." He has hired the high-powered Washington law firm of Akin, Gump, Strauss, Hauer & Feld to advise on how best to cope with Congressional interest.

Baltimore would like to see NIH put in place procedures that will "neutralize the activities of such as Mr Stewart and Dr Feder by quickly responding to charges of fraud and misconduct." But NIH's recent attempt to write new fraud guidelines has been stymied by the White House Office of Management and Budget.

The subcommittee has since gone a step further and borrowed Stewart and Feder from NIH on an as-needed basis to provide expert advice for future investigations. Stewart has already conducted interviews with investigators in Boston relating to the *Cell* paper.

Among those questioned were Henry Wortis and Brigitte Huber of Tufts University, both on the committee that had previously examined the paper. Both are concerned at the way a witness in an investigation has become an investigator.

NIH's attempts to investigate the Baltimore case have not run smoothly. It appointed a five-member investigatory panel early this year. But it came out before Congress that some of the members had worked with Baltimore, raising conflict-of-interest questions.

A new committee had to be appointed and the process started again, but finding five qualified people at short notice was difficult. Instead a three-member panel — Ursula Storb of the University of Chicago, Hugh McDavitt of Stanford University and Joseph Davie of Searle Pharmaceuticals — was appointed.

Their brief, according to Mary Miers who directs NIH misconduct investigations, is to see "whether the published paper was supported by the data...and to see what, if any, corrective action would be needed if there were errors".

The NIH committee has been given a very short time to draw its conclusions. NIH runs the risk that it will not fully satisfy Congress and invite further investigation. The Dingell subcommittee has received reports of more than a dozen new allegations of misconduct since April.

Joseph Palca & Alun Anderson

## Scientist for president

### London

THE appointment of Professor F. Bruno Straub as the new president of Hungary was last week recommended by the Central Committee of the Hungarian Socialist Workers' Party (HSWP). This is a significant step in implementing one of the major reforms advocated by the HSWP conference last month — the separation of the functions of state and party. For Straub, although a member of the Hungarian National Assembly (Parliament) since 1985, is not a member of the HSWP.

The party conference urged greater democratization of Hungarian political life, but ruled out the possibility of more than one political party. A greater role for the national assembly, and for its indepen-

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

*Straub—from enzymologist to president.*

dent members, is one of the recommended means of achieving this goal. Although the presidency is a ceremonial post, the nomination of a non-party man to this post is a big surprise.

Straub, an enzymologist, is a vice-president of the Hungarian Academy of Sciences. He is well-known internationally, having for many years represented the academy in the International Council of Scientific Unions. In the early 1970s, he was the prime mover in establishing the academy's Biological Centre at Szeged, and, as this was partially financed by United Nations funds, in organizing training courses for young biologists from developing countries. More recently, through the national assembly's environmental council, he has exerted considerable influence on policy decisions, including the protection of Lake Balaton, and the drawing up of measures intended to minimize the adverse effects of the Gabčíkovo–Nagy-maros hydroelectric project on the Danube.

Vera Rich



# NSF unwilling to count its chickens after last year's cuts

Washington

BOTH houses of Congress have now passed authorization bills for the National Science Foundation (NSF), and the House of Representatives has passed an appropriations bill, with the Senate soon to do the same. In normal years, this would make NSF confident about its 1989 budget. But after last year's last-minute cuts, nobody is counting on anything until the various money bills become law.

An authorization bill gives NSF permission to spend money in various ways, but it does not actually provide the cash. That comes from the appropriations bill. At the moment, the House version of the appropriations bill, passed last week on

the floor of the House, gives NSF \$1,885 million, about 10 per cent more than last year, but well below the 19 per cent increase that the administration is seeking. In the Senate, the appropriations bill has emerged from committee at \$1,880 million, with action on the floor expected soon.

Normally, with such close agreement between the two appropriations bills, NSF could expect a final funding level somewhere between the two figures. But last year NSF began to make plans based on similar early signs, only to find that all budgets changed at the last minute after a 'budget summit' between Congress and the White House fixed spending levels.

Congress does not accept all NSF's plans and seems unwilling to accept a request for full five-year funding for the new science and technology centres made in the president's 1989 budget. NSF had been seeking \$150 million for these centres, but the Senate has offered \$30 million for 1989 whereas the House only agrees to \$20 million.

The decision not to provide full funding for the centres put Congress in the unusual position of suddenly having an extra \$120–\$130 million to allocate without exceeding the president's budget request. The House and Senate took slightly different routes with this 'new' money. Both gave money for university facilities, but the Senate version contains provisions encouraging NSF to spend more on poorer universities, undergraduate education and university-industry cooperation.

The House version of the authorization contains a provision so noxious to the Reagan administration that it has threatened to veto the bill if it is included in the final version. The provision requires that a new ice-breaker being sought for NSF's Antarctic programme be built in the United States. Unfortunately, the only US bid is nearly double the \$100 million needed for a ship built overseas. House and Senate staff are now working over the two versions of the bill before a formal conference committee meeting.

With a presidential election coming up, NSF has joined the list of groups anxious to offer policy advice to a new administration. Frank Press, president of the National Academy of Sciences, has indicated his intention to set budget agendas that can be used by congressional committees in determining their spending priorities for science. Now the National Science Board, under its new chairman Mary Good, will also be picking topics on which it can offer well-informed opinions. The question is, as Good concedes, with all the voices clamouring to be heard, who will be doing the listening?

Joseph Palca

# Superconductor pessimism in US

Washington

THE failure, real or perceived, of American industry to engage fully their Japanese counterparts in the battle to develop commercial uses of the new high-temperature superconductors has contributed to a general perception of the inadequacy of the US research and development enterprise. This week the Congressional Office of Technology Assessment (OTA) has added its voice to the debate with a mostly pessimistic report\* on the relative standing of the United States and Japan.

Although the US government intends to spend \$95 million on superconductivity research during this fiscal year, most of this amount has been diverted from other programmes. Because superconductor development is necessarily multidisciplinary, involving chemistry, materials science and electronic engineering among others, redirecting funds from these areas may dilute the research effort.

Furthermore, the report notes, half of the superconductor money is going to the Department of Defense, whose track-record in commercially advantageous research is undistinguished, and another quarter is for the Department of Energy, whose forte is large projects with specific goals rather than small-scale, flexible research. OTA argues that better use would be made of federal funds if a greater proportion went to the National Science Foundation (NSF), which at present has been allocated only \$14.5 million for fiscal 1988.

The OTA report also repeats the familiar complaint that US industry is too short-sighted in its pursuit of technological development. A few large companies — AT&T, DuPont and IBM — are spending large amounts on superconductor research, but there are no US equivalents to the numerous and diverse Japanese manufacturers whose leaders evidently see long-term profits in superconductors. One remedy is for the US government to get directly involved in industrial research by sharing costs with industries willing to support basic research.

The report concludes that adherence to the traditional *laissez-faire* approach to technology transfer will concede the advantage to the Japanese, and argues that the US government must actively push the research priorities it deems essential. This might be done, it says, through the creation of a new federal agency devoted to civilian technology, to support research which is too speculative for industry but too "unglamorous" for the existing agencies.

David Lindley

\*Commercializing High-Temperature Superconductivity: OTA-ITE-388, June 1988.

# Ministers encouraged by EUREKA success

Paris

IF European industry is to present a united front in the international high-technology market, it must make unification of standards a priority from the outset of research and development. This was one of the main resolutions of the fifth ministerial meeting of the Commission of the European Communities' EUREKA high technology research and development programme, held in Copenhagen on 15–16 June.

EUREKA was started in 1985 as a non-military alternative to participation in the US Strategic Defense Initiative programme. Last week's approval of 54 new projects reflects the consistent annual increase in proposals from industry and brings the total number of active projects to 213, worth 4,756.2 million ECU (1 ECU = £0.69).

Ministers were able to see three projects near completion — the high-definition television (HDTV), a modular assembly-line applicable to different industries (FAMOS) and EUROTRAC, a project to monitor and deal with atmospheric and marine pollution. EUROTRAC is one of the few examples of a non-commercial application of EUREKA funds and, to be effective, has meant that Eastern bloc countries have been asked to take part.

French research minister Hubert Curien said that agreement on standards is essential if European industry is to compete with US and Japanese manufacturers. The "stupid" incompatibility of electricity mains frequencies between Europe, the United States and Japan is greatly complicating development of products, he said.

Having initiated EUREKA, France remains its most active supporter, with involvement in 102 projects, an investment of 1,374.8 million ECU.

Peter Coles



## Romania takes a firm line on chemical waste

London

SENIOR Romanian officials have been disciplined over the illegal dumping of chemical waste. Several people have been dismissed, including Foreign Trade Minister Ilie Vaduva, Stefan Birlea, the chairman of the State Planning Committee, and a Secretary of State at the Foreign Trade Ministry, Constantin Stanca. Three deputy prime ministers, the Minister of the Interior, the Minister of Finance and the Minister of the Petroleum Industry have received public reprimands or warnings. The director of the free port of Sulina and the Chimica Bucharest Company, and several executive officials of these organizations have been dismissed and now face prosecution.

The waste, described officially as "chemical and petrochemical residues", which are unofficially reported to have contained polychlorinated biphenyls, was transported to Romania in three shiploads of 800 tonnes each by Italian vessels. The contract covering the shipment was, however, made with a company registered in Liechtenstein, Chimica Liechtenstein. The import of hazardous material without proper safeguards is forbidden under Romanian law. Nevertheless, according to the official Romanian media, the administration of the free port of Sulina at the mouth of the Danube signed a contract in 1986 with Chimica Liechtenstein entitling the latter to import and deposit hazardous wastes at Sulina, thereby "abusing the special status" of the free port.

In 1987, the Romanian Foreign Trade company Chimica Bucharest also concluded a contract with the Liechtenstein company, with the intention, it is said, of "utilizing" the wastes. The officials concerned were either actively concerned in the deal, or else failed to notice what was going on, or, if they did notice it, failed to take appropriate measures.

The matter is being dealt with at the highest possible level — the dismissals were first announced during a session of the executive committee of the Romanian Communist Party, chaired by General Secretary Nicolae Ceaucescu in person. Neither Ceaucescu nor the Romanian authorities are noted for their concern over environmental matters. During the past few months, there have been protests from inhabitants of Ruse, just inside Bulgaria, about chemical fumes pouring in from Romania. Indeed, Romanian media commenting on the dumping affair concentrate more on the illegality (and implied corruption) than on the actual or potential danger.

Vera Rich

## Genentech gets patent protection for tissue plasminogen activator

Washington

GENENTECH, the only US company approved to sell the blood-clot dissolving drug, tissue plasminogen activator (TPA), now has patent protection for its product. Genentech holds the exclusive licence to the US patent granted last week to the Belgian technology company Innovi, patent agent for the University of Leuven.

The new patent is based on work done by Désiré Collen, at the University of Leuven. The patent is reported to be a broad "composition of matter" patent covering purified TPA itself, whether it is produced by recombinant means or by isolation from normal tissue.

Based on the patent's claims, Genentech has filed suits in the United States against both Burroughs Wellcome — the US division of the UK company Wellcome plc — and the Massachusetts biotechnology company Genetics Institute. The two companies are collaborating to produce their own version of TPA, and are expected to request approval some time this year to sell it in the United States.

Genentech is also currently embroiled in three other lawsuits over TPA. It has been sued by Abbott Laboratories in the United States for infringing a patent held by Abbott on a class of blood-clot dissolvers, and its Japanese patent is also being hotly contested by several companies in Japan (*Nature* 333, 587; 1988).

Meanwhile, the company's appeal against the decision that its UK patent is invalid is being heard in the UK high court. Genentech will be arguing its case for reinstatement of the patent, which was successfully challenged by Wellcome last year, and ruled to be too broad to stand.

The three appeal court judges are being advised on scientific matters by Dr Sydney Brenner, director of the MRC Molecular Genetics Unit in Cambridge. A decision is not likely for several months and, whatever, it is likely to be appealed.

The new patent granted to Innovi does not preclude the US patent office from granting additional patents for TPA. Companies, including Genentech, are seeking other, perhaps more specific, patents covering various forms of TPA and processes for making it.

Genentech will have to fight hard to establish and maintain patents that will give it sufficient means to stave off competitors developing TPA products. Numerous companies are working on "second-generation" forms of TPA that are improvements on the naturally occurring molecule, and it will not be clear what degree of protection the Innovi patent gives to Genentech until more infringement cases are settled in court. The effect on Genentech of the patent granted to the University of Oxford several weeks ago covering purified natural TPA with various sugar sidechains is also still uncertain (*Nature* 333, 383; 1988).

But Genentech is still the only company in the United States with approval to sell TPA, and its heaviest competition now comes from streptokinase, which is sold by Hoechst-Roussel and SmithKline Beckman for one-tenth of the \$2,200 price that Genentech charges for a single dose of TPA. To counter claims that its price is unfair, Genentech announced last week uninsured heart-attack patients will not be charged for the drug if they earn less than \$25,000 per year.

Carol Ezzell

## Human Frontiers gets a mention in Toronto

Tokyo

JAPANESE government officials in charge of the Human Frontiers Science Program are delighted that the programme was mentioned, albeit briefly, in the economic declaration of the summit of Western nations held last week in Toronto.

Japanese Prime Minister Noboru Takeshita brought up Frontiers at a working dinner on 20 June. He said that "until now Japan has imported the results of basic research, but from now on we want to export".

He pledged to follow in the footsteps of former Prime Minister Yasuhiro Nakasone and continue to promote the programme, and asked for the cooperation of the other summit nations. "The whole world will benefit from the promotion of science", he said.

It was touch-and-go whether Frontiers would be mentioned in the declaration, as economic and agricultural issues dominated the summit and environmental problems stole the scientific limelight. But in the end it was included under "other issues". The declaration noted the successful conclusion of the feasibility study of the programme, thanked Japan for providing summit nation scientists with the opportunity to contribute to the study and looked forward to the Japanese government's proposal for implementation of the programme "in the near future".

The declaration will provide a powerful bargaining tool when the Science and Technology Agency and Ministry of International Trade and Industry negotiate with the Ministry of Finance for funds for Frontiers later this year. David Swinbanks

# Embryo research ban causes ructions in West Germany

## Munich

A BAN on embryo research that would make West Germany one of the most strictly regulated countries in the world is moving closer to realization.

Although the stringency of the ban and the vehemence of its supporters may be peculiar to West Germany, the ethical issues raised by the law — such as its compatibility with the legality of abortion or the question of at what stage an embryo should begin to be shielded by law — are generally applicable.

The Deutsche Forschungsgemeinschaft (DFG) and the Max Planck Society (MPG) continue to oppose a ban, on the grounds that it would close the door on the possibility of therapies for terminal diseases (see *Nature* 327, 6; 1987).

Nevertheless, the MPG issued a statement on 27 June pledging to "support lawmakers in the process of clarification of the ethical and legal problems" involved. MPG president Heinz Staab promised a "strict prohibition" on any research involving human embryos "capable of development" even in early stages "until the bounds of ethically and legally justifiable" research are established.

The DFG said on the same day that it had never funded a project involving human embryos. Such projects are subject to voluntary guidelines of the Bundesärztekammer (a physicians' professional organization), which has reported no applications pending for embryo research.

An 'Embryo Protection Act' will be introduced into the Bundestag this year by the West German Justice Ministry. The law will prohibit research on human embryos under any circumstances, said Detlev von Bülow of the Justice Ministry.

The law would consider a fertilized human egg cell, from the point of the fusion of sperm and egg nuclei, as "deserving of legal protection". The law also prohibits research on totipotent cells removed at an early enough stage to be able to develop into a human being.

The new law would forbid the creation of human-animal chimaeras, the cloning of human beings from totipotent cells or the manipulation of germline cells — all of which are specifically opposed by the MPG and other scientific organizations.

Also to be banned is the creation of human embryos purely for the purposes of research, even for "high-level scientific goals" such as the improvement of *in vitro* fertilization (IVF) techniques or the potential use of fetal tissue or secretions for therapeutic purposes.

IVF as such will not be banned, but researchers would be forbidden to carry out experiments on or even observations

of embryos created for the purpose of implantation if there is even the slightest chance of damaging the embryo. Researchers at the seminar estimated that up to 1,000 children have been born in West Germany using IVF since the techniques became available about five years ago.

Violations would be punishable by up to three years of imprisonment and fines of an undetermined amount. It is this "criminalization" of research that the MPG and DFG oppose. Bundestag member Wolf-Michael Catenhusen (Social Democrat), the chairman of the Research and Technology Committee, mentioned the possibility of amending the Justice



Catenhusen — hinting at loopholes.

Ministry proposal to allow so-called "loopholes" for important research projects.

Seminar participants touched on several ethical questions that underlie the embryo research issue. The strict definition of an embryo was based, said Catenhusen, on the Nuremberg Codex, developed in the wake of the Nazi war-crimes trials, which forbids any experimentation on human beings without the consent of the experimental subject. The moment of fusion was chosen as the dividing line, he said, because, beginning with that moment, a "unique genome arises, which contains all the hereditary material necessary to create a potential human being". Other definitions have been used in other countries (see right).

Researchers who pointed out a logical contradiction in banning embryo research in a country that allows up to 200,000 legal abortions annually were rebuffed by von Bülow and others with two arguments. First, abortion is allowed only under conditions of "social emergency." Second, von Bülow cited a strict ban on embryo research as one way to make people think again about the rights of the embryo in the case of abortion.

Steven Dickman

# International outlook for embryo research

## Munich

THE international scene presents a varied picture regarding laws regulating research on human embryos. Albin Eser, a director of the Max Planck Institute for Foreign and International Criminal Law in Freiburg, West Germany, provided most of the information for this roundup.

West Germany is unusual in trying to pass a law concerning embryo research before such research has been done in the country. In the United Kingdom and Australia, the countries where much of the pathfinding research on *in vitro* fertilization (IVF) was done, medical researchers were allowed to regulate themselves for some years before legislation was considered. The state of Victoria in Australia did not pass a law until 1984, several years after research with human embryos had begun. The United Kingdom promised a law in 1985 but the issue is due to come before Parliament only this year. Many other countries are considering legislation, Canada, Switzerland, Austria and Sweden among them.

A ban on any type of research on embryos except 'research' which is done for the embryo's benefit seems to be unique to West Germany. Legislation in Australia forbids research on embryos under threat of fines or prison terms except where it has been approved by a central licensing authority. Voluntary guidelines in Switzerland and Great Britain have the same format.

But the situation in both of these countries could change when legislation is passed. Much depends on the definition of an embryo and the stage at which it merits protection under the law.

Countries such as France, which declared in a proposed law in 1984 that the embryo is a "legal subject" from the moment of fertilization, may nevertheless allow some research. Regulations in both France and Great Britain allow embryos to live on in culture past the time when implantation into a woman's uterus is normally carried out. The embryos only live for eight or nine days.

Another point of contention is what may be done with frozen embryos, which can be maintained practically indefinitely in a state of suspended animation. Switzerland's voluntary guidelines forbid freezing any embryos at all, even in the context of IVF therapy. The proposed West German law would not forbid the freezing of 'embryos' up to the point of sperm-egg fusion, but the proposed ban on research would render this procedure superfluous for any end but IVF. Britain and Sweden have tended to allow freezing for later research.

Steven Dickman



# Struggle ahead over abolition of academic tenure in Britain

London

ACADEMICS in the United States have expressed their support for their British colleagues in their fight against the legislation proposed by the government which would allow senior academics to be dismissed and replaced with junior staff purely for economic reasons. At the 74th annual meeting of the Association of American University Professors (AAUP) in Washington DC, a motion was passed urging the British government to withdraw this proposal and the proposal to abolish the system of tenure in universities, which, say the AAUP, has an "ultimate relationship" with academic freedom. The AAUP "deplore" the government's disregard for that relationship.

As the Education Reform Bill reaches its report stage in the House of Lords this week the British government is not expected to back down on the plans to abolish tenure. But neither is it likely to oppose the amendment successfully tabled by the opposition which gives statutory protection to academic freedom (*Nature* 333, 289; 1988). The Secretary of State for Education and Science is reported to be able "to live with" the

amendment.

This amendment, combined with the government amendments to set up grievance and disciplinary procedures, has satisfied the Committee of Vice Chancellors and Principals (CVCP) that academics will retain academic freedom.

The main matter which remains to be resolved this week is the terms of redundancy. Under the present form of the bill, and in accordance with the Governments expressed intention, "expensive" university academics may be made redundant and be replaced by a "cheaper" person to do the same work. This puts those staff at a disadvantage compared with all other employees in the country under the Employment Protection Consolidation Act (1978), including academic staff in other institutions of higher education. The CVCP considers this "grossly unfair."

Opposition has come from all parties in the House of Lords. The Conservative peer, Lord Beloff, is leading an amendment which would not allow redundancy under these terms. Another amendment proposes that the redundancy provisions be brought into line with the Act.

Another amendment likely to cause some debate in the Lords' two-day sitting with the bill is the government's amendment which gives university commissioners the power to draw up disciplinary procedures to deal with complaints relating to the appointment or employment of academic staff. An amendment is being tabled by the opposition to the effect that the onus would be on the university to prove there is good reason for dismissal before taking action. The amendment is backed by the Association of University Teachers, which represents more than 31,000 rank-and-file academics. The AUT considers this more important than allowing academics to use grievance procedures to redress unfairness. In that type of procedure the academic can receive compensation but is rarely reinstated, said the AUT.

Another problem with the bill which may cause the government some difficulty is one they appear to have brought upon themselves: abolition of tenure could quite possibly upset their programme of restructuring in university departments. Under the present form of the bill, staff with tenure lose it only if they change position. So staff are unlikely to agree willingly to reshuffling when this would bring loss of tenure. An amendment has been tabled which would mean that if academics moved in compliance with restructuring plans they should be exempt from loss of tenure.

That would have far-reaching effects as

departments in line for review include physics, chemistry, biology, dentistry and accounting. A review of Earth sciences departments was recently completed. The final cost of the full restructuring programme may be more than the government expected. Last week university vice chancellors asked for another £50 million to aid the process. The government has already allocated £155 million over a period of three years to the universities for this purpose.

Christine McGourty

## German education takes up a lifetime

Munich

New data from the Wissenschaftsrat, West Germany's science advisory council, reveal that in the natural sciences and engineering, a German university education may take the better part of a lifetime. The council released similar results for the social sciences and humanities at the end of April.

In engineering, students completed a



master's course in an average of 13 university terms, three terms longer than the universities recommend. The figures are complete to the end of 1985.

Government student loans run out after 10 terms. Only 10 per cent of engineering students completed their studies in the recommended time. The average age of graduates was 28 for master's students and 33 for students with doctoral degrees.

In the natural sciences, the results (to the end of 1985) were much the same. But wide variations showed up among the different universities. In Constance, for instance, physics and biology students completed their studies in an average of fewer than 10 university terms. In Berlin's Free University, by contrast, students completed a master's in physics in 14.6 terms and in biology in 14.2 terms. At the Technical University of Berlin, a physics master's degree took 16.6 terms, the highest average of all universities for physics.

Only 5 per cent of physics graduates were women compared with 47 per cent of biology graduates.

Steven Dickman

## AIDS vaccine trials to begin in Geneva

Paris

THE Swiss drug company Ciba-Geigy AG has announced imminent plans to start human safety trials of a potential vaccine against the AIDS virus. According to Pablo Valenzuela, deputy director of research at the California-based company, Chiron Corporation, which is a joint partner in the project, the vaccine uses part of the virus envelope protein, GP120, with "a special adjuvant". The decision to test the preparation in man follows pre-clinical animal trials on several species in which a strong immunological response was apparently observed.

The vaccine is to be tested on 25 healthy male volunteers at a Geneva hospital, part of Ciba-Geigy's network of Swiss contacts for clinical trials. Valenzuela says that similar trials are scheduled to begin in the United States "before the end of the year". The so-called "phase 1 safety trials" aim to see whether the vaccine produces antibodies to HIV (human immunodeficiency virus) in man and at what levels. Although GP120 has been shown to elicit an immunological response in sub-human primates, there is still no evidence that it provides effective protection against HIV infection.

Peter Coles



# Postal ban proposed on exchange of dangerous microbes

Washington

RESPONDING to pressure brought by Jeremy Rifkin and his Foundation on Economic Trends, the US Postal Service is proposing to ban sending disease-causing microbial agents and toxins through the mail. The ban would mean that some 100 microbes that are now exchanged between research institutions

under special packaging requirements will have to be sent by other means.

The Assistant Postmaster General last week announced the new policy at a Congressional hearing on the dangers of sending biological materials through the mail.

Regulations have been on the books since 1980 to ensure that dangerous microbes — such as the bacterium which causes anthrax, and the Crimean Congo haemorrhagic fever virus — are shipped in proper packaging. Rules set down by the US Department of Health and Human Services (HHS) currently require that vials containing such microbes be placed into three successively larger containers along with enough packing material to absorb the culture if the vial should break. The outermost container must carry a universal biohazard symbol and a hotline number at the US Centers for Disease Control (CDC). The most dangerous cultures must be sent by registered mail so that the sender receives confirmation that the package has arrived safely. The Department of Transportation also has authority over the interstate shipping of hazardous cultures of over 50 ml.

The ban proposed by the postal service would cover the specific list compiled by HHS of some 100 microbes that are known to cause human disease. Some are responsible for dreaded diseases such as Ebola fever or plague, but others are routinely exchanged in many areas of scientific research. All species of mycoplasma and mycobacteria are on the list, as are hepatitis and herpesviruses, simian virus 40, and all serotypes of *Escherichia coli* which can cause intestinal disorders.

The CDC and the American Type Culture Collection (ATCC), the US repository for all isolated microbial strains and genetic material, are the two most frequent mailers of biologically hazardous materials. CDC mailed about 1,000 disease-causing cultures over the past 17 months, and the ATCC reportedly shipped 900. Ft Detrick, Maryland, the centre of the Army's biological defence programme, claims to have sent only 48 cultures of hazardous microbes through overnight private courier services.

## NASA panel seeks more life in space

Washington

PRESIDENT Reagan's call for a permanent US presence in space has rekindled commitment to space exploration and made particularly timely a new report\* by a National Aeronautics and Space Administration (NASA) advisory committee. The Life Sciences Strategic Planning Study Committee (LSSPSC) emphasizes the importance of life sciences to the success of future manned space flights.

The report contends that the success of life sciences depends on quick and easy access to space, something a space station should provide. A space station would also permit long-term experiments to fill gaps in current knowledge about human survival in outer space.

To broaden the expanse of NASA's knowledge in sustained space operations, the committee also recommends cooperating with the Soviet Union on experiments, as well as encouraging universities to participate more actively in NASA's programme.

The LSSPSC report complements an earlier report† by the NASA Center Science Assessment Team (CSAT) which singled out life sciences as one of two speciality areas of science on which NASA should expend more time and effort. The CSAT report contends that a shortcoming of the NASA programme was a lack of long-term goals in the life sciences division. The LSSPSC report attempts to rectify that.

Although there is a fairly large body of data on the effects of prolonged spaceflight on human physiology and biochemistry, these data have not been well integrated into a theoretical framework. The LSSPSC report also indicates a need to pursue exobiological experimentation. A better understanding of the origin of life on this planet could well come from the exploration of how life evolved, or failed to evolve, on other planets.

Lisa L. Lyles

\*Exploring the living universe: a strategy for space life science. A report of the NASA Life Sciences Strategic Planning Study Committee. Washington, DC, 1988.

†Science at the NASA field centers. A report of the NASA center science assessment team. Washington, DC, 1988.

## Network for Europe

London

THE European network for research on the neural mechanisms of learning and memory has got off the ground with the first meeting of the coordinating committee. The network, proposed at the European Science Foundation general assembly last November, aims to provide a general forum where European workers in this research area can discuss work in progress with the goal of encouraging collaborative and interdisciplinary research projects.

European research in learning and memory has been suffering from poor exchange of information and ideas in a field that uses techniques ranging from biochemistry to animal learning and computer modelling. As a first move to fill this gap, the committee will set up a register of laboratories working in the field and organize several small specialist workshops. The initial phase of the network, lasting two years, has a budget of FFr 560,000 to fund these workshops and one open meeting, which will be held in the spring of 1990, probably in London. The committee will also promote links between experimentalists and modellers through liaison with programmes, such as the artificial-intelligence-based BRAIN project.

The emphasis will be on several well-established experimental preparations, to prevent too broad a spread of resources, but regular discussion meetings will ensure that new developments are not overlooked and results are seen in a wider context. If the first phase is successful, the network will be extended, and suitable projects may receive twinning grants.

Information about the network can be obtained from the secretary, Professor Jean Delacour, Université de Paris VII, Laboratoire de Psychophysologie, 7 quai St Bernard, Bât.B, 5,ét., F-75251 Paris Cedex 05, France. Jennifer Altman

The CDC report that no problems have been encountered with the current policy governing the shipment of disease-causing microbes. The CDC hotline averages only 50 calls per year reporting damaged packages with biohazard labels, and only three per year turn out to be actually leaking. No one has so far developed disease from a leaking package.

But what most concerns Congress is that no federal agency now monitors whether or not the packaging rules for disease-causing microbes are being followed regularly. At the Congressional hearing last week, a postal employee from the district in Maryland which handles mail to and from Ft Detrick reported several instances where improperly packed cultures opened during normal handling: one was even alarmingly marked "retrovirus". Carol Ezzell

## DFG re-elects Markl

Munich

THE members of the Deutsche Forschungsgemeinschaft (DFG) re-elected president Hubert Markl almost unanimously to a second three-year term last week at a DFG's annual meeting in Bonn. The term runs from January 1989 to December 1991.

Steven Dickman

## Effects of global change

SIR—Interest in environmental changes of global scale is increasing dramatically—in the scientific community, in agencies of government and among the general public. And in their planning, business and industry are beginning to expand their horizons in ways that would not have been considered a few years ago.

An earlier wave of environmental concern in the 1960s and 1970s prompted many corporations to add in-house staff or to contract with external companies to carry out environmental studies. In many cases, the resulting reports were of questionable objectivity and completeness, and of limited usefulness. The questions posed were often too narrowly circumscribed, and answers were produced only sporadically and were received at too low an administrative level to be effective.

The situation has changed in fundamental ways. New sources of data and new techniques of analysis have led to exciting scientific insights in a variety of fields. Cause and effect are being linked more closely as understanding of the physical and biological systems and their interactions increases. Time horizons have been greatly extended. For example, numerical results using climate models suggest that increasing atmospheric CO<sub>2</sub> due to burning of coal, oil and gas may change global distributions of temperature and rainfall in coming decades with accompanying changes in agriculture and rise in sea level. Radical proposals have been advanced to counteract the projected increase in CO<sub>2</sub>. As another example, evidence that the "Antarctic ozone hole" has resulted from release of certain chlorine compounds led last September to the Montreal Protocol, by which 38 countries have agreed to reduce production of the offending compounds by 50 per cent by the year 2000. Other examples concern clearing of tropical forests, extinction of animal and plant species and so on.

Consequences of large-scale environmental changes are beginning to affect policy decisions of business and industry more profoundly than ever before. A new wildlands policy introduced by the World Bank in 1986 states that "failure to preserve natural capital by wildland conservation in the present greatly increases the capital costs of economic development in the future" (S.G. Fitzgerald, *Bioscience*, **36**, 712–715; 1986). In accordance with this policy, the bank will weigh ecological considerations in deciding on future loans for hydropower development, roads, logging and other projects affecting wildlands. It will need to consider the probabilities of large-scale environmental changes which might affect each of its projects. Many corporations also will find their strategies and tactics influenced

significantly by the prospect of large-scale environmental changes.

For example, timber companies planning reforestation for harvests 50 years in the future must now consider the effects of increasing atmospheric CO<sub>2</sub> on climate and tree growth. Although current climate models provide only limited and fragmented understanding of climate change and climate variability, if used wisely they could be helpful in reaching decisions on where replanting should occur and on types of trees to be planted.

Companies engaged in coastal land development must take account of the projected rise in sea level and the resulting flooding of lowlands, increased coastal erosion, saline intrusion into estuaries, inundation of wetlands and so on. Model predictions might be used to project future rises in sea level and long-term changes in coastal processes. It would clearly be unwise to rely simply on model predictions—uncertainties are considerable—but neither should the implications of climate models be overlooked.

How can a corporation insure that its decision makers are sensitive to the implications of large-scale environmental changes? And how can it acquire the capability to respond appropriately to further new research results and new insights? Simple reliance on an in-house staff or on contract studies is likely to be ineffective for the reasons stated earlier. A better format could utilize an advisory group, composed of individuals of maturity, integrity and broad perspective. Members should have effective contacts with relevant research, including interdisciplinary research. Individual perspectives should be refined and further developed through regular internal discussions and through briefings. Such a body might utilize work carried out by an in-house staff or by contract, but its chief function would be to ensure that major corporate decisions were made with appropriate consideration of the effects of large-scale environmental changes. Few corporations provide this quality and depth of expertise at policy levels. There no known blueprints; institutional innovation may be called for.

ROBERT G. FLEAGLE

*Department of Atmospheric Sciences,  
University of Washington,  
Seattle, Washington 98195, USA*

## Who exploits whom?

SIR—J.A. Barnett (*Nature* **332**, 106; 1988) presents a singularly one-sided assessment of the financial returns and scientific satisfaction to be obtained by authors and editors of multi-author books.

The role of the editor and the amount of

work he or she puts into a particular book will clearly vary. At the outset there is usually protracted correspondence between the editor and the proposed contributors. For the book I am currently editing, this has been spread over a period of about four months. Once the contracts have been issued, one then has a breathing space of about a year before the manuscripts begin to arrive. (It is, however, worth bearing in mind that the editor himself may well be writing a chapter.) The amount of editorial work then required is very much determined by the quality of the manuscripts and whether or not they have been prepared according to the instructions to authors. On occasions I have spent ages correcting a weighty manuscript, even then to be forced into rejecting it as being unsuitable (only, I may say, after much heart-searching discussion with members of the editorial advisory board and correspondence with the author concerned).

Other problems are caused by authors who back out at the last minute and those who request a lengthy extension to the submission date, thereby throwing the whole production schedule into chaos.

The idea that the editor obtains an unreasonably large amount of money from the publisher, and that authors get much less than a fair deal, is, I believe, incorrect. In fact nobody does very well from these books. Both the scientific publisher and the editor have an important role to play, along with the author, in disseminating knowledge, which in the form of the authoritative and up-to-date review chapter can be of great value to the scientific community.

J. R. HARRIS

*North East Thames Regional  
Transfusion Centre,  
Crescent Drive, Brentwood,  
Essex CM15 8DP, UK*

## Inconsistency

SIR—The irony of your leader attacking the reductionist programme in molecular biology (*Nature* **333**, 11; 1988) was not lost on those of us who remember some other leaders. Poor Isidore Nabi was banished by you to a reductionist purgatory (*Nature* **293**, 2; 1981) and Sheldrake's book was burned by you (*Nature* **293**, 245; 1981) for criticisms of reductionist orthodoxy far less stringent than yours. Where can this inconsistency lead? A galilean recantation? A leader on the silliness of particle physics? Reductionists, like good schoolmen, crave belief in the proper order of things: a clockwork heaven and a reliable *Nature*. Don't abandon them in their hour of need.

ROGER BRADBURY

*Marine Systems Analysis Group,  
Australian Institute of Marine Science,  
Townsville, Queensland 4810, Australia*



# Can a Greek tragedy be avoided?

*Now under way is the third formal investigation within two years of a paper from the Whitehead Institute. In the fever of investigation, scientific reputations may be damaged.*

THERE are the makings of a Greek tragedy in the protracted inquiry into the conduct of an experiment reported more than two years ago from the Whitehead Institute (see page 788). Already, two enquiries have been carried out. A committee appointed by the National Institutes of Health is due to report in the near future, while the House of Representatives oversight committee (Rep. John Dingell, chairman) has yet to declare its intentions in detail.

The potentially tragic elements are plain. The research whose authenticity has been questioned is part of a stream of experimentation widely admired for its imaginativeness and productivity, and which offers the particular promise of an understanding of the regulation of the immune response in mammals in the language of molecular biology, not just in the phenomenological language of immunoassays.

Those caught up in the investigations include distinguished people of high reputation, not least Dr David Baltimore, at 51 one of the most respected members of the US scientific community (as much for his co-chairmanship of the National Academy of Sciences committee on AIDS as for his Nobel prize for the discovery of reverse transcriptase). The danger in this corrosive business is that nobody's reputation will be unscathed.

## Criticism

Criticism has come from two sources — internally, from Dr Margot O'Toole, a postdoctoral worker with Dr Imanishi-Kari at the Whitehead Institute, who told the Dingell committee two months ago of her failure to confirm experimentally data published in 1986, and from Dr Ned Feder and Mr Walter Stewart, two NIH scientists who have become well known (some would say notorious) for their crusade over the past several years on behalf of integrity in experimental science.

Feder and Stewart's activities have been much resented on several grounds, partly because they have no substantial scientific published record of their own, partly because they are self-appointed keepers of the scientific conscience and partly because of what often seems their nit-picking persistence. But in Stewart's long relationship with this journal (Feder is also a long-standing family friend) going back at least to 1970, the worst that could have been said of them is that they write at

too great a length. Although O'Toole has left research, at least for the time being, she as well as Feder and Stewart could be hurt by what lies ahead, even if their criticisms turn out to be correct.

The tale is that of a Greek rather than some other kind of tragedy because there are reasons for believing that it is unnecessary, and that it might even now be exorcised.

What follows first is an account of the disputed experiments.

The disputed article (Weaver, D. *et al. Cell* **45**, 247; 1986) reports a study of the production of immunoglobulins of different class and specificity by the lymphoid cells of one of several strains of transgenic mice founded two years earlier by David Baltimore and Frank Costantini (Grosschedl, R., Baltimore, D., Weaver, D. & Costantini, F. *Cell* **38**, 647; 1984).

At the outset, the authors may have thought this a simple system in which to study the mechanisms by which the immunoglobulin genes are rearranged and expressed in different kinds of cells. The immune system is notoriously complicated, as it must be to be capable of mounting an immune response to virtually any antigen. For example, although it is now possible to describe how the gene segments are rearranged so as to give the 'variable' regions of the heavy chains of immunoglobulin genes their specific properties, the regulation of this process remains unclear.

The transgenic mice are derived from early embryos of the strain C57BL/6, into which are injected copies of a gene encoding the heavy chain of an immunoglobulin molecule. A crucial feature of the design is that the gene introduced is derived from a different strain of mice (BALB/c) and that it is a functionally rearranged gene in which the appropriate variable genetic components corresponding to the variable region of the immunoglobulin molecule have already been stitched onto the sequence coding for the constant region, in this case a  $\mu$  gene (the characteristic constituent of IgM).

The formation of the rearranged heavy-chain gene was originally provoked by immunizing BALB/c mice with the chemical (4-hydroxy-3-nitro-phenyl) acetyl, NP for short. This means that the protein corresponding to the heavy chain, when associated with an appropriate light chain (either the  $\kappa$  or the  $\lambda$  versions are

possible) will be an antibody specific against NP.

Apparently, the injected copies of the heavy-chain gene are incorporated into the genome of recipient mice at places unconnected with the regular immunoglobulin genes. In the 1984 experiments, transgenic mice were found to have been 30 and 130 copies of the heavy-chain gene. The transgenes were expressed in substantial amounts only in lymphoid tissues (B and T cells, which are also present in spleen, thymus and lymph nodes).

## Objective

The essential objective of the experiments described in 1986 was to study the effect of the presence of transgenes on the production of other kinds of immunoglobulin molecules. The central, and surprising, conclusion of the 1986 article is that the range of antibodies produced by cells expressing the endogenous antibody genes is profoundly affected by the presence of the transgene. In particular, the transgenic mice produce a large proportion of antibodies that share with the exogenous (transgene) antibody the property known as idiotype.

Idiotype is a concept that has helped give immunology a bad name, at least in part because it is more easily defined in general terms than measured. But in principle, the idiotype of an antibody refers to a structural feature of the variable regions of the antibody that itself provokes the formation of other antibodies. Operationally, an idiotype is defined by the class of other antibodies with which it reacts uniquely. Structurally, it will eventually be possible to say in which elements of the structure of an immunoglobulin molecule the property resides.

As Weaver *et al.* say, there is a "profound alteration of the antibody repertoire . . . endogenous immunoglobulin products very frequently mimic the idiotype, but not generally the [larger] polypeptide structure, of the transgene product".

This conclusion is taken (at least by Imanishi-Kari), to be support for the network theory of the immune response due to Dr Niels Jerne (he, too, was honoured by a Nobel prize) based on the idea that the immune response is either amplified or suppressed by the recognition by lymphocytes of the idiotypes of immunoglobulin molecules on the



surfaces of B cells. (David Baltimore is on record as holding that other explanations are more probable.)

The idiotype of an immunoglobulin molecule is defined by the assays in which it is measured. Starting with an antibody specific against, say NP, the first step is to make antibodies against that molecule by using it to immunize other animals. Even among mouse antibodies specific against NP, not all will react in the same way against the various anti-antibody antibodies produced by immunization.

Imanishi-Kari and her associates, over the past half-dozen years, have been able to classify the antibody molecules specific against NP produced by C57BL/6 mice empirically defining six classes of idiotypic reactions for the latter. For the 1986 study, a reagent for identifying immunoglobulin molecules was made by immunizing guinea pigs with a stretch of polypeptide spanning the rearranged variable region of the immunoglobulin heavy-chain gene.

### Transgenic mice

The 1986 study describes a series of immunological and nucleic acid studies to pin down the characteristics of the immunoglobulin molecules produced from the tissues of transgenic mice. In practice, Weaver *et al.* uncontroversially followed Grosschedl *et al.* in using DNA analysis of blood from the tail of offspring mice to tell which carried the transgene.

Part of the confusion that has arisen with the controversy of the past year arises because critics of the 1986 article (O'Toole and Feder and Stewart) attack two (or perhaps three) of the immunological assays used in the study, but its authors insist that their main conclusion is correct. In logic, both camps could be correct; some of the results of the assays could have been in error, but the main conclusion could have been sustained by other measurements. But Feder and Stewart's argument does not easily accommodate such a possibility.

On one intermediate conclusion, there seems nevertheless to be no controversy. The 1986 article says that the serum of transgenic mice contains amounts of antibody specific against NP which are also idiotypically identical (by the guinea-pig antibody test) with the transgene. Both  $\lambda$  and  $\kappa$  light chains are involved in the construction of these antibody molecules. The amounts of antibody carrying transgenic idiotype in the serum are up to 70 times greater than in the control mice (for this purpose, the littermates found not to be carrying the transgene).

The production of large amounts of a specific antibody by mice never challenged by the antigen concerned may be explained by a suggestion in the 1984 article that the regulation of the immune

response in B cells consists of regulation of the process of rearrangement; once genes have been rearranged, and if they are suitably equipped with means of initiation, mRNA will be transcribed from them and peptides produced. Naturally, that requires production of light chain from the endogenous light-chain genes.

Controversy centres on the techniques for telling whether the antibody molecules in the serum of transgenic mice are derived from transgenes or from endogenous genes. Plainly the diagnosis must be based on some part of the heavy chain other than the variable region. The distinction was made possible by the use of a distinctive constant region present in the transgenic but not in the endogenous gene.

This is a  $\mu$  gene sequence, which is known to differ from the endogenous  $\mu$  gene fragment of the recipient mice by a single amino acid. (Technically, the exogenous and endogenous antibodies are of different allotype.) The transgene has type  $\mu^a$ , the endogenous molecule of type  $\mu^b$ . Most of the experiments reported in the 1986 article bear on this question of telling whether the immunoglobulins in transgenic mice are exogenous (from the transgene) or endogenous.

The experiments are as follows:

(1) To differentiate between the endogenous and exogenous  $\mu$  chains in serum antibodies, the authors used two monoclonal antibodies; one of these called Bet-1, was believed to be specific for  $\mu^a$  and the other (called AF6) to be similarly specific for  $\mu^b$ . The specificity of these reagents has been questioned.

(2) To provide a general assessment of the composition of the serum antibodies carrying the idiotype in question, a guinea-pig anti-idiotypic antibody was used. The polypeptide encoded by the transgene is more strongly bound than are the serum antibodies in the mice, suggesting that the serum antibodies cannot be largely composed of molecules encoded by the transgene. This assay is not disputed.

(3) Further to clarify this conclusion, the authors made hybridomas from B cells taken from both spleen and lymph nodes of transgenic mice, again using the Bet-1 and AF6 reagents to tell whether the  $\mu$  chains were endogenous or exogenous. Two conclusions stand out: the proportion of hybridomas using the  $\lambda$  light chain is much greater than the usual 1–2 per cent, while a substantial proportion (1 out of 43 and 10 out of 129) of the hybridomas had used endogenous  $\mu$  chains in antibody molecules. The first of these conclusions is not disputed, but the second is.

(4) As a more direct test for the presence of antibody encoded by the transgene, the group tested the hybridomas for the presence of RNA transcribed from the transgene. Such RNA molecules can be protected from attack by nuclease by

means of a piece of DNA corresponding to the known genetic structure of the transgene. The conclusion is that most mRNAs tested are *not* protected, suggesting that their structure is not that of the transgene. The authors also carried out a series of hybridization experiments (Northern blots) which they claim provide support for endogenous  $\mu$  production. These experiments are disputed by O'Toole on the grounds that these diagnostic techniques are less sensitive than the immunological assays. The authors say that it is especially significant that two hybridomas from which the transgenic  $\mu$  gene had been lost were nevertheless producing similar molecules, presumably by rearrangement of endogenous genes.

(5) The nucleotide sequences of gene fragments spanning the variable region in hybridoma molecules have been determined for cDNA derived from mRNA gathered from hybridoma cells. There appear to be many differences between hybridoma cDNA and the DNA sequence of the transgene. Moreover, the sequence analysis also shows that the constant regions of many of these genes are often of the  $\gamma$  rather than the  $\mu$  isotype. These conclusions are not disputed.

O'Toole seems to have been the first to raise doubts about the validity of the paper's chief findings, which were formally put forward in a memorandum dated 6 June 1986, and which forms part of her evidence to the Dingell committee.

Briefly, she then asserted that the Bet-1 reagent is not an unambiguous test between  $\mu^a$  and  $\mu^b$ . This view is based on O'Toole's own work with the reagent in Imanishi-Kari's laboratory.

### Objection

This objection is echoed by Feder and Stewart, using the data recorded in the 17 laboratory notebook pages in O'Toole's possession. They are first concerned with the criterion used for scoring the assays for idiotype expression and transgene expression.

Their objection, which applies to the hybridoma data (experiment (3)) particularly, is that the decision to score as "positive" those assays in which a well on a plastic sheet yielded more than 1,000 decays (counts) per minute in an immunoassay machine, and all the others as "negative", took no account of the fact that the data, when graphed, appear to be the tail of a distribution, introducing an arbitrary element into the classification scheme.

They also echo O'Toole in remarking that, when some of the hybridomas (still in culture) were retested a few days after the first measurements, they were found to be more powerful sources of antibody than on the first occasion, suggesting that the scoring of antibodies as "positive" or

"negative" would have been exquisitely sensitive to the passage of time.

A further elaboration of this argument, based again on the 17 pages of laboratory notes, is that the test data for the hybridomas derived from transgenic spleen tissue in reality show a good correlation between measurements with the anti-idiotypic reagent and Bet-1, but with a slope suggesting that the latter is only 30 per cent as sensitive as the latter. Feder and Stewart, however, say that both reagents must be "looking at" the same protein, which they suppose must be the transgene.

A bizarre twist to this argument by Feder and Stewart came to light earlier this week. The two have previously argued that the data for hybridomas based on lymph-gland tissue must have been scrambled: one of the grounds for this assertion is that no fewer than 8 of 10 hybridomas scored positive against AF6 (the reagent for the endogenous  $\mu$  gene) were immediately next to each other in the records, suggesting a technical error of some kind.

Feder and Stewart now claim to have unscrambled the data they allege were "scrambled". They assert that the laboratory notebook pages in question each had 52 lines for data entries, and that on the first page (of three) dealing with the lymph data, the idiotypic data have been entered correctly but the Bet-1 data (counts per minute) begin with what should have been the 63rd record. "Unscrambling" the data according to this rule leads them to the conclusion that the Bet-1 and idiotypic data are again well-correlated, with a slope not very different (0.28 rather than 0.30) from that obtained with the data from lymph-node hybridomas.

## Critics

Feder and Stewart's critics will quickly point out that it is not a valid basis for a criticism of other people's work to replace their data set by another obtained by rearranging the set of one variable in a pair. Their defence will no doubt be that the rearrangement is exceedingly simple, and that it could have arisen inadvertently by cutting up a printed tape, of the kind that carries the output from radio-immunoassay machines. But this argument is at best circumstantial.

There is criticism both from O'Toole and Feder and Stewart of the fact that the controls used for the hybridoma assays reported in the 1986 paper had not been examined at the same time as the other mice. The explanation is that the mouse selected as a control turned out, after the event, to have been a transgenic and not a pre-strain mouse. (Curiously, the 17 pages of laboratory notes also turned up among the records of the breeding colonies of transgenic mice.)

O'Toole also says that normal mice more often produce NP-specific antibodies than is recorded in the published paper, and differs from the authors in their conclusion that most of the idiotype antibodies are endogenous on the grounds that a "completely inadequate assay was used to show that exogenous  $\mu$ " was not expressed in transgenic hybridomas".

These are only some, if apparently the more substantial, of the criticisms that have been levelled at the 1986 publication.

There is a sense in which the disputed paper does its case much less than justice. Thus in the categorization of hybridomas formed from transgenic mice, Weaver *et al.* record substantial numbers carrying the NP-idiotypic (43 out of 150 spleen hybridomas, for example), but then literally do not refer to the significance of the fact that only 10 of these were IgM in type, and that somehow the transgene had stimulated the production of IgG and other immunoglobins all carrying the idiotype marker. The article does however draw attention to the important circumstance that some of the results can be explained only if parts of the  $\mu$  transgene segment have been rearranged within the cells in which they find themselves.

The fact that those close to this field are persuaded by the data, whatever their faults, that there is something distinctly odd about the transgenic mice, seems to be part of the reason why so much tension and personal animosity has been engendered by these events.

It is also important that, while none of the critics have used the word "fraud" (and that Feder and Stewart in their congressional testimony said that they were talking only of "error"), many of the charges made against the 1986 article carry that imputation. Error may be forgivable, but denial of the possibility of error when charged with it may be less easily forgiven.

That may explain why so much contention has been engendered by this otherwise humdrum affair. Both the accusers and the accused have regarded each other with suspicion, even disdain. One consequence is that their discussions have not been successful. Maybe they should look for a simpler way. (Baltimore, evidently keen to get Feder and Stewart off his back, suggested to NIH last year an inquiry that would bind all concerned to accept its findings, and to secrecy thereafter; the first demand was reasonable, the second impracticable.)

The much more serious danger is that, while the participants in what promises to be an endless series of inquiries into their conduct decline to communicate with each other, the issue will indeed fall to be decided by some congressional committee, say Dingell's, and that the standards of propriety and improp-

riety will from then on be set by the US Congress?

The difficulty is peculiarly American. It is proper to respect "due process", but unwise (as the Founding Fathers of the United States knew well enough) to allow respect for law to inhibit honest conversation. The chances are that the 1986 article in the dispute is flawed in some of its analyses, but that it has also demonstrated that transgenic mice are an interesting way of learning how the immune system is regulated. That process will continue, uninhibited by the errors (alleged or otherwise) in the 1986 paper. Who, in ordinary circumstances, would complain of that, error (unproven) notwithstanding?

## Remedy

The remedy in this sad and sometimes shabby case does not rest with Ned Feder and Walter Stewart, with Margot O'Toole, David Baltimore or even, with great respect, John Dingell. It is a communal problem, that of how the scientific community should learn to live with the evident imperfections of the scientific literature. Leaving aside fraud, mere error is only half the struggle: what is to be done about mere mis-referencing? And is the prevalence of that not, by extension, a sign that other aspects of the literature are less than fair?

Instead, the remedy rests with the scientific community. People have been shy of speaking out about the Baltimore affair (as it is unfairly called). Should they not say what they believe to be the truth, without waiting to see whether they will lose friends in the process, or be sued for libel? It is a curious paradox that the country that, by the First Amendment and all that has flowed from it, should be hamstrung, for fear of what Congress might decide, from deciding for itself what constitutes respectable publication. Would the community or the Congress be able to live with a congressional undertaking to be the referee of last resort?

Mr Dingell, who is no slouch, would recognise the difficulty, and would shy from the responsibility. Recognizing that he has other things to do, he would find some way of shuffling off the responsibility for running the world's scientific journals (not all of which are published in the United States). So might not the scientific community calculate that Dingell will have learned that lesson in advance, that he and the rest of Congress will wish nothing but that researchers should look after their own business, freeing him (or them) for other causes? The snag is that, even in the United States, the community does not appreciate what it has to lose. And that the Dingell committee does not understand what damage it may do.

John Maddox



## Duchenne muscular dystrophy

# Localizing the gene product

Kurt G. Beam

Two reports on pages 861 and 863 of this issue by Arahata *et al.*<sup>1</sup> and by Watkins *et al.*<sup>2</sup>, as well as one by Zubrzycka-Gaarn *et al.* in an earlier issue<sup>3</sup>, demonstrate that the protein product of the Duchenne muscular dystrophy (DMD) locus is predominantly localized to the surface membrane of striated muscle cells. In another report on page 858 of this issue<sup>4</sup>, Chelly *et al.* describe the relative abundance in different tissues of the RNA transcript of the *DMD* gene. Understanding the tissue distribution and subcellular localization of *DMD* gene products provides a valuable starting point for understanding the pathophysiology of the disease and, it is hoped, for devising effective therapies.

Duchenne muscular dystrophy is a fatal, degenerative disease of muscle that afflicts about 1 of every 3,500 males born. The locus responsible for the disease was first identified on the basis of cytogenetic and genetic linkage studies. This locus is enormous, about 2,000 kilobases (kb) in size, the largest known for any human gene. A similar locus is also present in mice; 14-kb RNA transcripts of the mouse and human genes show an almost 90-per cent sequence similarity over the region for which they have been compared<sup>5</sup>. Polyclonal antibodies have been raised against peptides derived from complementary (c)DNAs for the 14-kb 'dystrophin' transcript, and these antibodies bind to a protein of relative molecular mass ( $M_r$ ) ~ 400,000 (400K) that is present in extracts of skeletal muscle from normal humans and mice. The protein is absent from muscle extracts of humans with Duchenne muscular dystrophy and mice with the X-linked, recessive mutation *mdx*<sup>6</sup>. Although it has become common usage to refer to the immunoreactive 400K protein as dystrophin, the identity between it and the dystrophin protein predicted from the cDNA sequence of the 14-kb transcript has yet to be established rigorously.

## Tissue distribution

The 400K protein is a minor component (by mass) of skeletal muscle, representing only about 0.002 per cent of its total protein content. The protein is also present in cardiac muscle, at a level comparable with skeletal muscle, and smooth (gut) muscle, at a lower level than skeletal muscle<sup>6</sup>. The tissue distribution of the 14-kb dystrophin RNA transcript is characterized by Chelly *et al.* in this issue<sup>4</sup>. Because the transcript is both very large and present in low abundance, the authors have developed a

highly sensitive assay based on the polymerase chain reaction. This assay starts with reverse transcriptase-mediated synthesis of single-strand cDNA from cellular messenger (m) RNA primed by dystrophin oligonucleotides. Dystrophin cDNA is then selectively amplified through a pair of synthetic primers as in the typical polymerase chain reaction. The system is calibrated using a parallel reaction primed for detecting aldolase A, whose mRNA is sufficiently abundant that its tissue distribution had been established earlier by more conventional means. Using this elegant methodology, Chelly *et al.* determine that the dystrophin transcript is present not only in muscle, but also in kidney, lung and brain (the abundance in these latter tissues being about 1 per cent of that in skeletal muscle). The presence in brain confirms earlier studies that used RNase protection assays<sup>7,8</sup>. These studies indicate that the absence of brain dystrophin accounts for the mental retardation that sometimes accompanies Duchenne muscular dystrophy.

The full-length cDNA encoding the human 14-kb transcript predicts a dystrophin protein whose amino-terminal domain is conserved with the actin-binding region of  $\alpha$ -actinin, and whose overall sequence suggests that it is a rod-shaped, cytoskeletal protein<sup>9</sup>. Support for a cytoskeletal function of dystrophin is now provided by the immunolocalization of what is apparently dystrophin to the surface membrane of heart and skeletal muscle<sup>1-3</sup>. Curiously, an earlier study<sup>10</sup> examining the relative abundance of dystrophin in subcellular fractions of homogenized muscle indicated that the protein was mainly associated with triads, the region of apposition between t-tubules and sarcoplasmic reticulum. Based on the newer immunocytochemical results that indicate a surface localization, one might imagine that the function of dystrophin is to help the muscle membrane to resist the stresses associated with contraction. In the absence of dystrophin, the muscle membrane would be subject to repeated, local tearing and the cumulative damage caused by this loss of cellular integrity would ultimately result in fibre necrosis. Indeed, this model, which was first put forward as the membrane hypothesis of muscular dystrophy<sup>11</sup>, fits well with the histopathological changes seen in dystrophic muscle.

Although the new results are in general agreement with a cytoskeletal function for dystrophin, they also raise several puzzles. Thus, three independent groups<sup>1-3</sup> all

agree that polyclonal antibodies, raised against peptides derived from the dystrophin cDNA, react with a protein that is associated with the surface membrane of normal muscle but is absent in DMD or *mdx* muscle. (Each of the groups used different peptides for antibody production than those used by the other two.) According to two of the groups<sup>2,3</sup>, the  $M_r$  of the immunoreactive protein is the expected ~ 400K. However, Arahata *et al.* in this issue<sup>1</sup> report a  $M_r$  of 210K (the presence of protease inhibitors should have prevented proteolysis). How is one to explain why the antibodies of Arahata *et al.* do not crossreact with the 400K protein? Perhaps the amino-acid sequence they selected for peptide synthesis is weakly immunogenic and is found in both the 210K and 400K protein products of the dystrophin transcript. Then, if the smaller of the proteins were more abundant, only it would be detected by Western blotting.

Is the 210K protein a product of the *DMD* locus? If it is, then one would expect that the antibodies used by Watkins *et al.*<sup>2</sup> and by Zubrzycka-Gaarn *et al.*<sup>3</sup> should have detected not only the 400K protein but also the 210K protein. If the 210K protein is the product of a different locus then it is necessary to explain why mutation of the *DMD* locus suppresses its expression in both humans and mice.

## Regeneration

Another puzzle is that the absence of dystrophin does not seem to be equally damaging to the different muscle cell types that normally express this protein. Based on the presence of the fetal isoform of the heavy chain of the muscle protein myosin as an indicator of fibre regeneration, Webster *et al.*<sup>12</sup> determined that type IIB (fast glycolytic) fibres of skeletal muscle are preferentially affected in Duchenne muscular dystrophy. Specifically, it seems that fast twitch fibres degenerate and subsequently regenerate much earlier, and to a much greater extent, than slow twitch fibres. With time, the ability of fibres to regenerate is lost as the pool of muscle precursor cells is exhausted. A simple prediction based on the observations of Webster *et al.*<sup>12</sup> is that

1. Arahata, K. *et al.* *Nature* **333**, 861-863 (1988).
2. Watkins, S.C., Hoffman, E.P., Slayter, H.S. & Kunkel, L.M. *Nature* **333**, 863-866 (1988).
3. Zubrzycka-Gaarn, E.E. *et al.* *Nature* **333**, 466-469 (1988).
4. Chelly, J., Kaplan, J.-C., Maire, P., Gauthron, S. & Kahn, A. *Nature* **333**, 858-860 (1988).
5. Hoffman, E.P., Monaco, A.P., Feener, C.C. & Kunkel, L.M. *Science* **238**, 347-350 (1987).
6. Hoffman, E.P., Brown, R.H. & Kunkel, L.M. *Cell* **51**, 919-928 (1987).
7. Chamberlain, J.S. *et al.* *Science* **239**, 1416-1418 (1988).
8. Nudel, U. *et al.* *Nature* **331**, 635-638 (1988).
9. Koenig, C., Monaco, A.P. & Kunkel, L.M. *Cell* **53**, 219-228 (1988).
10. Hoffmann, E.P., Knudson, C.M., Campbell, K.P. & Kunkel, L.M. *Nature* **330**, 754-758 (1987).
11. Mokri, B. & Engel, A.G. *Neurology* **25**, 1111-1120 (1975).
12. Webster, C., Silberstein, L., Hays, A.P. & Blau, H.M. *Cell* **52**, 503-513 (1988).

dystrophin is preferentially expressed in fast twitch fibres. But, at least for the cross-sections in the new reports<sup>1-3</sup>, all fibres display fairly equal staining for dystrophin. Furthermore, Fig. 1 of ref. 2 on page 864 shows that the quantity of dystrophin in extracts of soleus muscle (predominantly slow twitch) equals or exceeds the level in EDL muscle (predominantly fast twitch).

There are other instances in which cells express dystrophin but are comparatively unharmed by the absence of the protein. Thus, skeletal muscle from *mdx* mice lacks dystrophin<sup>1,2,6</sup> and shows histological changes similar to those observed in DMD muscle. Clinically, however, *mdx* mice differ little from their normal counterparts. Similarly, the absence of dystrophin from DMD cardiac muscle produces histological changes but cardiac function seems to be much less impaired than skeletal-muscle function. Certainly, tissue

and species differences can be invoked as an explanation for this differential susceptibility to damage (*mdx* mice may retain the ability to regenerate muscle fibres throughout their relatively short lifespan). Additionally, this differential susceptibility prompts the idea that there are important, unidentified proteins distinct from dystrophin and whose expression is influenced by the *DMD* gene locus.

Given the intense research activity directed towards understanding Duchenne muscular dystrophy, the inconsistencies are not surprising. Rather, it is encouraging that the results that are rapidly emerging from several laboratories are, in many ways, consistent. The results that do not fit with a tidy picture are of the sort that should stimulate profitable lines of investigation. □

Kurt G. Beam is in the Department of Physiology at Colorado State University, Fort Collins, Colorado 80523, USA.

## Particle physics

# Symmetry violated in kaon decay

David J. Miller

ANOTHER prediction of the standard model of weak interactions has been confirmed in an experiment performed by the NA31 collaboration at CERN (Burkhardt, H. *et al. Phys. Lett. B* **206**, 169–176; 1988). This has measured the ratio of the relative rates for long- and short-lived neutral kaon decays to two neutral pions and to two charged pions. Because this double ratio is slightly less than 1, Burkhardt *et al.* can exclude the 'superweak' theory which had been introduced specially to explain neutral kaon decays. The result is consistent with the existence of six quark flavours, with the mass of the top quark (so far unobserved) being between 50 and 100 GeV.

Kaons have always been useful probes of particle interactions. They have strangeness, the first of the heavy-flavour quantum numbers, which gave the first clue that nuclear matter is composed of quarks. Because the strange quark is heavier than the ordinary up and down quarks, strange particles decay to ordinary particles, converting the excess mass of the strange quark into the kinetic

energy of the decay products. Both charged and neutral kaons decay to several final states, including two- or three-pion ( $\pi$ ) states.

There was consternation in the 1950s when it was discovered that kaons could decay into either two or three pions. Particle interactions were assumed to preserve the symmetries of wavefunctions and decay amplitudes under various simple operations. The parity operation  $P$  reverses the spatial coordinates, equivalent to mirror reflection with a rotation through  $180^\circ$ . Charge conjugation  $C$  exchanges particles and antiparticles. The  $CP$  operation does both at the same time. Two-pion states have even parity — the sign of the wavefunction remains the same after the  $P$  operation. The three pions from kaon decay have odd parity — the sign reverses. A single particle has a wavefunction with well-defined parity, so the random decay of kaons into either two or three pions suggests that parity is not conserved in the decay process. This clue, confirmed in studies of nuclear  $\beta$ -decays, led to the modern theory

of weak interactions.

After parity violation had been proved, it was still expected that weak interactions would conserve the  $CP$  properties of the decaying particles. Sums (mixings) of neutral kaon and antikaon wavefunctions can be either odd or even under the  $CP$  operation (Fig. 1). Two-pion states are even under  $CP$  and the three-pion states from kaon or antikaon decay are odd. The rest mass of three pions takes up most of the available energy in a kaon decays (413 MeV out of the kaon rest mass of 498 MeV) so, if  $CP$  is conserved, the odd- $CP$  combination of kaon and antikaon should decay quite slowly, to three pions, compared with the even- $CP$  combination which should decay rapidly to two pions with a big release of energy. This explains the long-lived ( $K_L^0$ ) and short-lived ( $K_S^0$ )

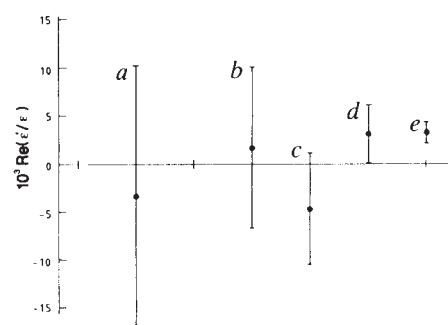


Fig. 2 Progress in measurements of the real part of  $\epsilon'/\epsilon$  since 1972. a, Holder *et al.* (1972); Banner *et al.* (1972) and Christenson *et al.* (1979); b, Black *et al.* (1985); c, Bernstein *et al.* (1985); d, Woods *et al.* (1988); e, CERN NA31 collaboration (1988).

components seen in neutral kaon decays.  $CP$  appeared to be conserved — until 1964 when J.H. Christenson *et al.* discovered that one  $K_L^0$  decay out of every 3,000 is to the  $CP$ -forbidden two-pion channel.

The 'superweak' theory (Wolfenstein, L. *Phys. Rev. Lett.* **13**, 562; 1964) explained the known facts in an *ad hoc* way, including the near equality of the two experimental ratios ( $\eta$ ) of decay rates, both zero if  $CP$  is conserved:

$$\eta_{+-} = \frac{K_L^0 \rightarrow \pi^+ \pi^-}{K_S^0 \rightarrow \pi^+ \pi^-}, \quad \eta_{00} = \frac{K_L^0 \rightarrow \pi^0 \pi^0}{K_S^0 \rightarrow \pi^0 \pi^0}$$

A better theory did not appear until M. Kobayashi and K. Maskawa (*Prog. theor. Phys.* **49**, 652; 1973) made a very bold prediction.

'Neutral current' weak interactions had just been discovered in neutrino interactions, but never seen in the strangeness-changing decays of kaons. A second heavy quark — with charm — had been suggested to account for this inconsistency, but it had not yet been observed. Kobayashi and Maskawa suggested another two heavy quarks, beyond charm, and showed that the normal theory of weak interactions would then be able to generate a small  $CP$ -violating amplitude in neutral-kaon decays.

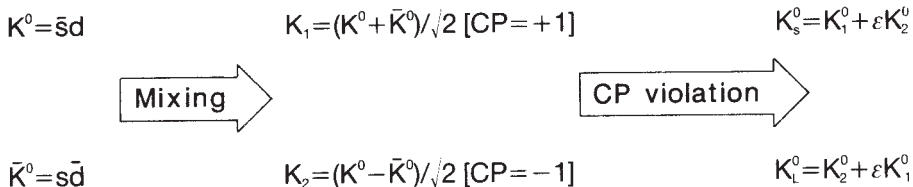


Fig. 1 The kaon ( $K^0$ ) and antikaon ( $\bar{K}^0$ ), made from strange ( $s$ ) and down ( $d$ ) quarks and antiquarks, undergo two levels of mixing to give the  $CP$ -violating forms  $K$ -short and  $K$ -long. Their decays to two pions depend both on the mixing ( $K$ -long would not decay to two pions at all otherwise) governed by the parameter  $\epsilon$  and a direct contribution governed by the parameter  $\epsilon'$ .



Now that both charm and beauty particles have been discovered, the Kobayashi-Maskawa theory has become an integral part of the standard model, although several of its predictions are still being tested. One was proved correct last year: the neutral beauty mesons and their antiparticles mix together just as the neutral kaons do (see my report in News and Views, *Nature* **329**, 394; 1987). The crucial test will be to produce particles containing the sixth quark — 'truth' or 'top'.

But the ratio of  $\eta_{00}$  to  $\eta_{+-}$  is also very important. Superweak theory predicts equality, with only a slight mixing of CP-odd in the  $K_S^0$  wavefunction and of CP-even in the  $K_L^0$  given by a mixing parameter  $\varepsilon$  which is known to be non-zero (see Fig. 1). Kobayashi-Maskawa theory predicts CP violation both in the wavefunctions and directly in the decay process, via a parameter  $\varepsilon'$  which is zero in superweak theory. The decay rates are modified so that  $\eta_{00} = \varepsilon - 2\varepsilon'$ ;  $\eta_{+-} = \varepsilon + \varepsilon'$ .

The problems of measuring a difference between  $\eta_{00}$  and  $\eta_{+-}$  are formidable. Each CP-violating decay of  $K_L^0$  to two pions is masked by 3,000 normal  $K_L^0$  decays. Experiments need large statistics, and extremely clean identification of the two-pion final states to reduce systematic errors caused by contamination from CP-conserving decays. The CERN NA31 collaboration (and a competing collaboration at Fermilab, Chicago; Woods, M. *et al.*, *Phys. Rev. Lett.* **60**, 1695; 1988) uses proton beams with hundreds of gigaelectron volts of energy to produce fast kaons whose decay products travel forward, close to the beam direction, inside a long vacuum tank with a detector at the end. At these high energies a single detector can measure charged pions and the  $\gamma$ -rays from neutral pions with good precision, and at the same time identify and reject electrons, muons and extra pions from the unwanted decays.

Using samples of one million  $K_L^0 \rightarrow \pi\pi$  and 10 million  $K_S^0 \rightarrow \pi\pi$  decays, the CERN group finds  $\eta_{00}/\eta_{+-} = 0.980 \pm 0.004$  (statistical)  $\pm 0.005$  (systematic); or the real part of  $\varepsilon'/\varepsilon = (3.3 \pm 1.1) \times 10^{-3}$  (with combined errors). This is only three standard deviations away from equality of  $\eta_{00}$  and  $\eta_{+-}$ , but in close agreement with predictions from Kobayashi-Maskawa theory. Figure 2 shows how the errors on the real part of  $\varepsilon'/\varepsilon$  have been reduced in successive experiments since 1972. Further improvements are still needed, of course, both to confirm this result and to measure the size of the phase angle in Kobayashi-Maskawa theory which fixes the strength of CP-violation and which is one of the basic parameters in the standard model.  $\square$

David J. Miller is in the Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.

## Ore deposits

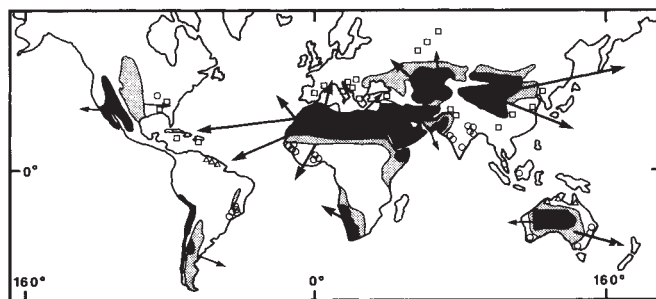
# Bauxites gathering dust

K. Pye

BAUXITE ores, consisting of oxides and oxyhydroxides of aluminium and iron mixed with variable amounts of clay, quartz and heavy minerals, provide the main source of aluminium and their genesis is therefore of considerable commercial interest. Three main modes of origin are traditionally proposed for bauxites and laterites: *in situ* weathering and leaching of bedrock to leave aluminium- and iron-rich residues; Al/Fe enrichment of sediments or weathering residues by movement of groundwater; or deposition

and petrographic data strongly suggest that wind-borne (aeolian) material has been introduced at the top of the profile. This is consistent with the continental dune trends and occurrence of a zone of kaolinite and quartz-rich sediments in the adjacent part of the Indian Ocean that indicate a dominant movement of sand and dust westwards and north-westwards across Western Australia.

The possibility that other bauxite deposits in Australia have also been significantly influenced by aeolian processes



Main dust trajectories (arrows) and location of principal bauxite deposits on  $\square$ , karst;  $\triangle$ , clastic sedimentary rocks;  $\circ$ , igneous and metamorphic rocks. Bold areas are arid; stippled are semi-arid.

of Al/Fe-rich sediment reworked from pre-existing bauxites or lateritic soils. These processes are not mutually exclusive, and some bauxites have complex histories involving all three<sup>1,2</sup>. On page 819 of this issue<sup>3</sup>, Brimhall *et al.* present the first substantive evidence, based on work in the Darling Downs area of Western Australia, that deposition of wind-blown dust can also contribute to metal enrichment in bauxites. These findings will aid the interpretation of bauxites elsewhere, and provide further evidence that the transport of wind-blown dust is a process of geological, climatological and economic significance.

There are about 900 million tonnes of bauxite reserves in the Darling Downs area<sup>4</sup>, representing almost a quarter of Australia's reserves and making it one of the largest deposits in the world. The bauxite near Jarrahdale contains a large amount of quartz sand and an abraded heavy mineral suite, suggesting that it developed by weathering of a former fluvial sedimentary cover<sup>5</sup>. Another interpretation<sup>6</sup> is that it formed by *in situ* weathering of the underlying Archaean bedrock, although this idea does not account for the abraded heavy mineral suite.

The more rigorous chemical mass-balance approach adopted by Brimhall *et al.*<sup>3</sup> in their study in this issue demonstrates that *in situ* weathering cannot fully account for the concentration of elements in these profiles, and their mineralogical

warrants investigation. The Mitchell Plateau deposits lie beneath the north-westerly dust trajectory, and the deposits in Victoria, northern New South Wales and southern Queensland lie in the path of a south-easterly dust trajectory towards the Tasman Sea (see figure). The important deposits at

Weipa on Cape York Peninsula and at Gove in the Northern Territory do not lie downwind of an obvious dust source, although some deflation and westwards reworking of lateritic soils on Cape York could have occurred. North Queensland and the Northern Territory today experience a seasonally wet tropical climate, but during the Pleistocene<sup>6</sup> conditions were much drier for long periods, with dust blowing possibly on a larger scale.

The main dust trajectories are apparently correlated with the location of bauxite deposits in other parts of the world too (see figure). The deposits in Guinea, Sierra Leone and Ghana are in the area affected by dust trajectories from the Sahara. Deposition of iron-rich Harmattan dust could have significantly influenced the composition of soils in

1. Valetton, I. *Bauxites* (Elsevier, Amsterdam, 1972).
2. Bardossy, G. *Karst Bauxites* (Elsevier, Amsterdam, 1982).
3. Brimhall, G.H. *et al.* *Nature* **333**, 819-824 (1988).
4. Sadleir, S.B. & Gilkes, R.J. *J. geol. Soc. Aust.* **23**, 333-344 (1976).
5. Grubb, P.L.C. *Econ. Geol.* **66**, 1005-1016 (1971).
6. Kershaw, A.P. *Nature* **272**, 159-161 (1978).
7. Vine, H. in *Desert Sediments Ancient and Modern* (eds Frostick, L.E. & Reid, I.) 171 (Blackwell, Oxford, 1987).
8. Macleod, D.A. *J. Soil Sci.* **31**, 125-136 (1980).
9. Rapp, A. *Geol. Föhr. Stockh. Föhr.* **105**, 161-168 (1984).
10. Prospero, J.M. *et al.* *Earth planet. Sci. Lett.* **9**, 287 (1970).
11. Prospero, J.M. *et al.* *Nature* **289**, 570-572 (1981).
12. Blumel, W.D. *Catena Suppl.* **1**, 67-95 (1982).
13. Watson, A. *Sedimentology* **32**, 855-876 (1985).
14. Whalley, W.B. in *Chemical Sediments and Geomorphology* (eds Goudie, A.S. & Pye, K.) 197-226 (Academic, London, 1983).
15. Walker, T.R., Waugh, B. & Crone, A.J. *Geol. Soc. Am. Bull.* **89**, 19-32 (1978).

northern Nigeria<sup>7</sup>. Similarly, the karst bauxites of southern Europe occur in an area influenced by northerly dust transport from the Sahara. There is increasing evidence that widespread *terra rossa* soils in southern Europe contain a significant amount of weathered dust<sup>8,9</sup> which has been locally reworked.

Transport of Saharan dust as far as the Caribbean and South America has also been documented<sup>10,11</sup>, and it is possible that the parent material for the Jamaican karst bauxites was partly aeolian. Fine-grained dust, which is transported great distances, is typically enriched in aluminium and iron relative to silicon because of a high content of clay minerals and amorphous aluminosilicate material

compared with the coarser local dust. Furthermore, the relatively low rate of dust accumulation at great distance from the dust source could favour further aluminium and iron enrichment by leaching of silicon in suitable topographic traps.

The importance of aeolian dust in the formation of other near-surface residua and precipitates, including some calcretes<sup>12</sup>, gypcretes<sup>13</sup>, desert varnish<sup>14</sup> and red beds<sup>15</sup>, is now well-established. Following the new work of Brimhall *et al.*, bauxites need to be re-evaluated to assess the importance of dust deposition and aeolian reworking in their genesis. □

K. Pye is in the Department of Earth Sciences, University of Cambridge, CB2 3EQ, UK.

## Reverse genetics

# Anti-sense RNA as a tool to study plant gene expression

Conrad Lichtenstein

ANTI-SENSE RNA, regulatory RNA that can inhibit gene expression, is complementary to a given messenger RNA species, with which it pairs, forming double-stranded RNA. In one way or another (see later) this duplex RNA formation blocks expression of the protein gene product. Anti-sense RNA is a natural mechanism for gene regulation in bacteria but has also been used experimentally (see ref. 1 for a review). On page 866 of this issue<sup>2</sup>, van der Krol and co-workers report on the use of such RNA to block gene expression in plants in a dramatic and beautiful way. By insertion of an anti-sense gene suppressing expression of an enzyme required for pigment synthesis, they generate transgenic plants with altered flower colour.

With the advent of recombinant DNA technology, genetic engineering meant introducing new and different genes into organisms. A later development was to manipulate the existing genes within an organism. One way to achieve this 'gene surgery' exploits existing cellular machinery for genetic exchange by homologous recombination. Here, a cloned DNA sequence is mutated *in vitro* and then reintroduced back into a cell where it recombines with and so replaces a target gene, providing that there is a reasonable stretch of common DNA sequence.

This reverse-genetics technology is a powerful tool to elucidate the genetic function of an unknown piece of DNA. It also has evident applications to gene therapy, for removing undesirable genes or replacing defective ones and for protein engineering. Although the principle has been successfully applied in the study of bacteria and yeast, it is less well-suited to plants and animals where transforming

DNA integrates randomly into non-homologous sequences rather more frequently than into the homologous target sites. Much research is devoted to overcoming this problem; various strategies allow a direct selection for homologous recombination<sup>3</sup> or screen to weed out the random from the targeted events in a population of genetically transformed cells<sup>4</sup>.

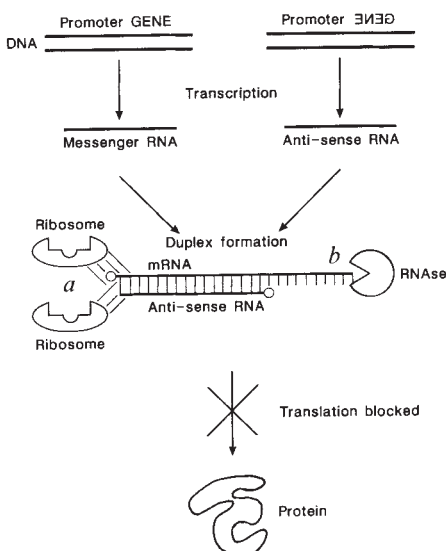
Alternative approaches to manipulating gene expression, using anti-sense RNA, are concurrently being developed (see figure). Here, anti-sense genes are constructed *in vitro* by reversing the orientation of a portion of transcribed DNA, usually including the region encoding the protein, and placing this next to a transcriptional-control sequence, a promoter. This gene cassette is then delivered to the

target cell, resulting in genetic transformation and inhibition of expression of the target gene. Thus, unlike homologous recombination, anti-sense RNA does not allow gene replacement, only gene turn-off. As this may not be completely effective, it is a possible disadvantage, but could be useful when a reduction in gene expression is desired. It also provides the potential for selective control of gene expression in specific tissues or at specific stages in development by an appropriately regulated promoter. Its chief advantage is that it can be used against invading nucleic acids to engineer virus resistance; clearly homologous recombination cannot be used in these circumstances.

How does anti-sense RNA regulate gene expression? Much can be learned by studying bacteria which have, not surprisingly, already invented the use of anti-sense RNA<sup>1</sup>. In bacteria one likely mechanism is by blocking translation. Evidence for this is seen, for example, in an experiment in which resistance to the RNA coliphage SP was engineered into *Escherichia coli*<sup>5</sup>. The gene for the maturation protein was found to be the best target for anti-sense RNA with this phage. The greatest inhibitory effect was observed with a sequence encompassing the 24-base non-coding region plus a 216-base coding region of the sequence. Significantly, even a 19-base sequence covering only the ribosome-binding site exerted a strong inhibitory effect on phage proliferation, suggesting that the presumed duplex RNA which forms in this region prevents ribosome recognition and hence initiation of translation into the protein. In the natural examples in bacteria it is significant that the anti-sense gene also spans this ribosome-binding-site region.

From micro-injection studies of anti-sense RNA into unfertilized eggs of frogs it can be seen that only sequences including the region spanning the site of initiation of translation are effective at reducing translation of the gene product, suggesting a mechanism similar to that in prokaryotes<sup>6</sup>. In other studies, where anti-sense genes were stably incorporated into the nuclear genome of cultured mammalian cells, there is evidence that an anti-sense RNA:messenger RNA duplex is formed within the nucleus but cannot be transported to the cytoplasm and hence is not available for translation. In this latter experiment, the region spanning translation initiation was not required and other regions were equally effective<sup>7</sup>. Other evidence suggests that the anti-sense RNA: messenger RNA duplex which forms in the nucleus is highly unstable and rapidly degraded<sup>8</sup>.

The first published demonstration of the use of anti-sense RNA-mediated suppression of gene expression in plants was using a transient expression system, electroporation, of plant protoplasts with



a, Prevention of initiation of translation; b, RNA degradation.



both sense and anti-sense constructs<sup>9</sup>. Later, gene suppression of a foreign gene was achieved in stably transformed plant lines following a second transformation with the corresponding anti-sense gene<sup>10</sup>. A problem with transformed plants, however, is the large variation in expression of foreign genes from one transformed line to another (reviewed in ref. 11).

Van der Krol *et al.* in their paper in this issue<sup>2</sup> overcome this problem by targeting their anti-sense gene against an endogenous plant gene encoding a petal-specific chalcone synthase, an enzyme that allows the development of red petals in, among others, *Petunia hybrida*. Blocking expression of chalcone synthase produces white or variegated petals resulting from variations, within petal tissue, in chalcone synthase expression (see figure on page 867). An analysis of the transgenic plants, which have one or more copies of the anti-chalcone synthase gene, shows heritable but variable alterations of petal pigmentation correlating with alterations

in both chalcone synthase enzyme and messenger RNA levels in petal tissue. The mechanism remains elusive, but evidence points to increased degradation of messenger RNA rather than inhibition of translation. This, along with other studies, should lead to a search both for the enzymes responsible for degradation of duplex RNA and for a natural role in higher organisms for gene regulation by anti-sense RNA. □

1. Green, P. *et al.* *Rev. Biochem.* **55**, 569–597 (1986).
2. Van der Krol, A.R. *et al.* *Nature* **333**, 866–869 (1988).
3. Smith, A. & Berg, P. *Cold Spring Harb. Symp. quant. Biol.* **49**, 171 (1984).
4. Smithies, O. *et al.* *Nature* **317**, 230 (1985).
5. Coleman, J. *et al.* *Nature* **315**, 601–603 (1985).
6. Melton, D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 144 (1985).
7. Kim, S. & Wold, B. *Cell* **42**, 129 (1985).
8. Crowley, T.E. *et al.* *Cell* **43**, 633–641 (1985).
9. Ecker, J.R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 5372 (1986).
10. Rothstein, S.J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 8439–8443 (1987).
11. Lichtenstein, C.P. & Fuller, S.L. *Genetic Engineering* (ed. Rigby, P.W.J.) Vol. 6, 103 (Academic, London, 1987).

Conrad Lichtenstein is in the Centre for Biotechnology, Imperial College of Science and Technology, London SW7 2AZ, UK.

## Mass extinctions

# Gradual volcanic catastrophes?

K. G. Cox

ON pages 841 and 843 of this issue, Duncan and Pyle<sup>1</sup> and Courtillot *et al.*<sup>2</sup> present new <sup>40</sup>Ar/<sup>39</sup>Ar dates for basaltic lavas from the Deccan trap continental flood basalt province of India, and together provide the most accurate information yet available for the age and duration of an immense volcanic episode. This is of exceptional interest as it confirms previous suggestions<sup>3</sup> that the province developed very rapidly and coincided closely with the Cretaceous/Tertiary boundary about 65 million years ago (Myr), at which important faunal and floral changes occurred. The new data must be woven into the continuing debate about whether faunal extinctions are triggered by volcanism or by meteorite impacts.

The Deccan traps province is far from unique. Since the late Palaeozoic (about 300 Myr) the Earth has been subjected to numerous similar volcanic episodes consisting of huge outpourings of basalt lavas onto continental surfaces. Among these, the Siberian traps (Permo/Triassic, 240 Myr), Paraná basalts (late Cretaceous, 130 Myr), and Karoo Province (early Jurassic, 180 Myr) all probably covered substantially larger areas than the Deccan<sup>4</sup>, and in the Karoo, volcanic sequences are locally very much thicker than anything so far recorded from the Deccan<sup>5</sup>. So the contention of Courtillot *et al.*<sup>2</sup> that the Deccan "could have been the largest volcanic event in Mesozoic and Cenozoic times" is hard to sustain.

Given that numerous volcanic episodes

have occurred, if we are to believe in a direct connection between mass extinctions and volcanism then a very significant correlation between the ages of the two phenomena must be demonstrated. There have been attempts<sup>6</sup>, but I suspect that the accuracy of dating in several important provinces (Siberia, Paraná, Karoo, Antarctic) is not yet sufficient to produce convincing results, and this leaves aside the parallel question of how accurately mass extinctions can be quantified and dated. Obviously a great effort from geochronologists, volcanic stratigraphers, palaeomagneticians and palaeontologists will be required to resolve such questions.

But the new data in this issue<sup>1,2</sup> on the duration, as opposed to age, of the Deccan volcanism raise another interesting point. Although a large province seems to have been erupted remarkably quickly on a geological timescale, it actually consists of several individual eruptive events separated by repose periods. The environmental effects of each eruption would undoubtedly be dramatic<sup>7</sup>, but the capacity of the biosphere to recover during repose periods would also be strong. Thus the creation of irreversible and permanent faunal or floral changes is likely to be highly dependent not just on the size of the province but also on the eruptive rate and the length of repose periods.

A crucial question then is how many eruptive events constitute a province? This is more difficult to answer than might at first appear. An eruptive event can be

defined as a period of continuous magma emission, in the cases under study mainly from fissures from which floods of basaltic magma emerge and travel for perhaps hundreds of kilometres. But the advancing flow front can cool and act as a dam whose frozen crust must be breached by magma closer to the vent. Flow thus proceeds by leap-frogging, and the geologist examining sections through the volcanic pile could find it extremely difficult to decide which of the flow units observed really represent separate events. Often, only the presence of a weathering horizon, in many cases a laterite, or an intercalation of sediment, gives unequivocal evidence of genuine repose. Detailed field and geochemical work can potentially elucidate such problems, but there are very few parts of the world where reliable information is available.

From field studies in the Deccan, I estimate that the number of eruptive events is probably quite small, perhaps 100–500. If the whole episode lasted for 500,000 years, an average repose period would last about 1,000–5,000 years. On a human timescale this seems a long time: one small laterite horizon could be equivalent to the whole of human recorded history. Fortunately, there are no historic analogues of flood basalt volcanism, but as a result the capacity of the environment to recover from the effects of a single eruption during repose periods of this sort can only be guessed. The Laki fissure eruption in Iceland in 1783, following which starvation reduced the population by half, is the nearest comparison, but 200 years on the effects are no longer remarkable. Extrapolation from this eruption, which released 12 cubic kilometres of basalt, is difficult because a typical flood-basalt eruption in the past may have been two orders of magnitude larger.

One can conclude that if continental volcanism has anything to do with mass extinctions it is probably via a series of episodes of environmental stress. Unlike the meteorite-impact hypothesis in its simplest form, the volcanic model can be regarded as a stepwise catastrophe, and the geological evidence has been interpreted in this way<sup>8</sup>. Hybrid models<sup>6</sup>, linking the volcanism itself to meteorite impact, introduce interesting complexities into the debate, as do models which postulate multiple impacts from a comet shower<sup>9</sup>. □

1. Duncan, R.A. & Pyle, D.G. *Nature* **333**, 841–843 (1988).
2. Courtillot, V. *et al.* *Nature* **333**, 843–846 (1988).
3. Courtillot, V. *et al.* *Earth planet. Sci. Lett.* **80**, 361 (1986).
4. Macdougall, J.D. (ed.) *Continental Flood Basalts* (Riedel, Dordrecht, in the press).
5. Erlank, A.J. (ed.) *Spec. Pap. geol. Soc. S. Afr.* **13** (1984).
6. Rampino, M.R. & Stothers, R.B. *Trans. Am. geophys. Un.* **67**, 1247 (1986).
7. Officer, C.B. *et al.* *Nature* **326**, 143–149 (1987).
8. Hallam, A. *Science* **238**, 1237–1242 (1987).
9. Hut, P. *et al.* *Nature* **329**, 118–126 (1987).

K. G. Cox is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

## Visual perception

## Contours revealed by concealment

Oliver Braddick

MANY of the cells in our visual system are specialized for reporting the presence of edges or lines. The accepted view is that these are important features because they indicate the boundaries of objects and surfaces, and so define shapes that we can recognize. But it is a rare object whose shape is completely defined by visible contours; in our cluttered world, the shape will usually be partly occluded by nearer objects. This means that some of the edges bounding a region in the field of view physically belong to it, but others are occluding contours, giving a fortuitous overall shape to the region. Another way in which occlusion affects visual processing is that the interruption of an occluded contour can provide visible evidence of the occluding edge, even when that edge is not revealed by any consistent luminance difference (Fig. 1). Findings on both these aspects of visual occlusion were reported at a recent meeting\*, and they suggest that detection of occlusion relationships occurs remarkably early in visual processing.

Cells which respond to edge or line orientation are numerous in both areas V1



**Fig. 1** Aligned line terminators, in a pattern which could occur where one surface occludes another, create an anomalous contour that does not correspond to any consistent luminance difference but which can activate orientation-selective cells in area V2 of monkey cortex.

and V2 of the primate cortex, the first two stages in cortical visual processing. R. Von der Heydt and E. Peterhans (University Hospital, Zurich) recorded the responses of these cells when stimulated with conventional luminance edges, and with patterns like Fig. 1 in which human viewers report 'anomalous' or 'subjective' contours. Cells in area V1 show only the responses that would be expected from their sensitivity to the orientation of the individual line segments. But many V2 cells show a response when the subjective

contour is oriented at the same angle as the optimal luminance edge for activating the cell — in Fig. 1 this would mean that the line segments were actually orthogonal to the cell's preferred orientation.

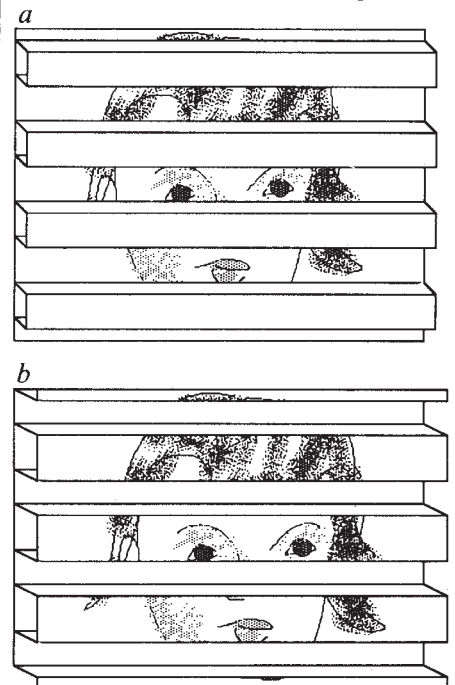
As with the perceptual effect, these responses depend strongly on the number of line segments defining the anomalous contour. It does not seem possible to explain this behaviour in terms of any simple summation of light and dark areas within a receptive field. Von der Heydt and Peterhans suggest that these V2 cells have two sets of inputs, one generating their response to light/dark edges or bars, and another combining signals from detectors which respond to suitably aligned terminators, of lines which are optimally nearly orthogonal to the orientation preferred by the V2 cell. This means that the V2 cell shows a stimulus equivalence for oriented contours which can be signalled by very different kinds of information in the image; either light/dark discontinuities, or the aligned terminators resulting from occlusion.

In the past, anomalous contours have sometimes been called cognitive contours, with the perception of an edge being regarded as a high-level inference from the sensory data. It is probably fruitless to argue where sensation ends and cognition begins. However, these results suggest that if there is such an inference, its outcome is present in the pattern of neural activity at a stage quite close to the input of the cortical circuits dealing with visual pattern information. The possible gap between animal neurophysiology and human perception is partly bridged by experiments<sup>1</sup> showing that cats can be trained to make a key-touch response to just those patterns which yield an anomalous contour in human perception.

K. Nakayama (Smith-Kettlewell Institute, San Francisco) argued that there are good reasons why occlusion information has to be made explicit early in visual processing. One set of line terminators in Fig. 1 would correspond to real physical features while the others are just a consequence of the occlusion relationship, and Nakayama's experiments show how that implies differences in almost every aspect of visual processing. Face pictures, for example, can be created with interrupting strips (Fig. 2). If these are presented stereoscopically, the bars can appear either in front or behind the plane of the picture. The latter is inconsistent with occlusion, and so the continuity of features between the separated strips is harder to perceive, leading to markedly

worse performance in subjects' recognition of familiar faces<sup>2</sup>.

There is a somewhat similar influence on motion perception. The direction of motion of a grating is ambiguous<sup>3</sup>, but the ambiguity can be resolved if the motion of the ends of the grating bars is visible. But Nakayama's group finds that this disambiguation effect depends on stereoscopically determined occlusion relationships: if the



**Fig. 2** The same strips of a face picture give different effects depending on whether the intervening bars appear in front, consistent with occlusion (a) or behind (b). Nakayama *et al.* use stereoscopic depth differences, which can be indicated here only by perspective.

grating appears in front of the surrounding frame, then the ends are seen as intrinsically belonging to the bars and therefore determine their motion, whereas if the grating is seen behind a window, the ends are an extrinsic consequence of occlusion, and their effect is much weaker<sup>4</sup>. The assimilation of colour and binocular rivalry are other effects that are altered by the apparent occlusion of surfaces.

Clearly, the problems of identifying and using occlusion information are embedded in many aspects of vision. What is less clear is how far the perception of anomalous contours can be identified with the perception of occlusion. In cases such as the well-known Kanisza triangle figure, the occlusion interpretation seems rather compelling, and indeed the anomalous contours disappear if they are given a stereo disparity which would force them to appear behind the figures they appear to occlude<sup>5</sup>. Discontinuities of the kind shown in Fig. 1, however, are not exclusively associated with one direction of occlusion, and might signal boundaries that had nothing at all to do with occlusion. Boundaries without luminance

\*Visual Processing of Form and Motion European Brain and Behaviour Society Workshop, Tübingen, 16–19 March 1988.



changes occur in other situations, for example where surfaces of different orientation meet at a corner but neither is concealed, or where two surfaces with different textures abut in the same plane. A mechanism of the general kind proposed by von der Heydt could also pick up alignments of local features for these purposes. The need to decide whether a contour is intrinsic to an object, or imposed by occlusion, could then arise just as strongly for an anomalous contour as for an ordinary luminance edge.

But the issue of whether anomalous contours are an input contributing to perceived occlusion or a consequence of it, rests on the presumption that visual

processing is primarily a one-way, bottom-up flow of information. Cognitive theorists have challenged this presumption for a long time, and as visual neuroscience starts to lay hands on their phenomena, they may wish to point to the wealth of descending pathways which the neuroscientists recognize in the anatomy but not much, yet, in their thinking. □

1. Bravo, M., Blake, R. & Morrison, S. *Vis. Res.* (in the press).
2. Nakayama, K., Shimojo, S. & Silverman, G.H. *Invest. ophthalm. Vis. Sci. (Suppl.)* **28**, 365 (1987).
3. Adelson, E.H. & Movshon, J.A. *Nature* **300**, 523–525 (1982).
4. Shimojo, S., Silverman, G. H. & Nakayama, K. *Invest. ophthalm. Vis. Sci. (Suppl.)* **29**, 264 (1988).
5. Blomfield, S. *Nature new Biol.* **246**, 256 (1973).

Oliver Braddick is Reader in Vision at the Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK.

## Molecular recognition

# Frontiers of the immune system

Charles A. Janeway Jr

SIR Winston Churchill, in hospital with an infection during World War II, is said to have looked at his hospital chart and asked his physician: "What are these lymphocytes?" He was told: "We don't know, Prime Minister." "Then why do you count them?" Churchill retorted. This anecdote illustrates how far immunology, and lymphocytes, have come since then. These small cells, with a homogeneous morphology of apparent inactivity, are the essential cellular component of the immune system, providing specific recognition of foreign antigen and initiating immune effector responses. The immune system is often compared with the nervous system: both detect and respond to alterations in the environment in a highly specific manner, both show memory, and both consist of complex organizations of cells having similar appearances but different functions. A chief distinction between these two biological systems is their very different anatomical organization: the nervous system has an observable and relatively static structure, whereas the immune system, and its lymphoid component in particular, is noted for its mobility. Furthermore, most immune responses can be mimicked using dissociated cells *in vitro*, which has tended to devalue studies of the anatomical organization of the immune system. However, this neglect of anatomy is waning, and papers like that of Goodman and Lefrançois on page 855 of this issue<sup>1</sup> will help this attitude to its deserved quietus.

Whatever Sir Winston's attitude to lymphocytes might have been, his attitude in 1940 is well known to have stressed comprehensive protection of the British Isles, with great emphasis on defence of the routes of invasion\*. The immune system has a similarly comprehensive brief in defending the body, and it

distributes its manifold forces in many distinct venues (see table).

The well-known humoral (or antibody-mediated) and cellular (or T-cell-mediated) effector mechanisms listed in the first two sections of the table combat infections with extracellular and intracellular pathogens, respectively, in the body as well as at epithelial surfaces through the transepithelial secretion of IgA. Despite the extensive distribution of these well-studied immunological defences, this picture is incomplete in two ways. First, the epithelial cells that make up the barrier between the internal and the external milieu are not themselves protected by the cells and molecules described above; and second, there is a cell type whose role in immune defences is not yet integrated into our thinking. During the search for the genes encoding the variable (V)  $\alpha$ - and  $\beta$ -chains of the T-cell receptor, Tonegawa and his colleagues unexpectedly discovered a third class of rearranging T-cell-specific genes that is similar to the  $\alpha$ - and  $\beta$ -receptor genes<sup>2,3</sup>.

## New gene family

The discovery of this gene family, now called  $\gamma$ , is an interesting example of how molecular genetics can drive biology. The discovery of the  $\gamma$ -gene has revealed a great deal about the immune system that not only was not previously known, but was not even suspected, most notably the existence of a third family or lineage of lymphocytes distinguished by the presence of a novel receptor identified using anti- $\gamma$  peptide antibodies comprising the variable  $\gamma$ - and  $\delta$ -chains<sup>4,5</sup>. Once it became clear that the  $\gamma$ -chain was associated with a second variable chain, called  $\delta$ , the genes encoding this fourth rearranging T-cell-specific gene were identified and characterized in the mouse<sup>6,7</sup> and in man<sup>8,10</sup>.

Having found this new receptor, the story moves from reverse genetics to reverse biology: the gene sequence was used to identify the protein, and the protein was used to identify the cells in which the gene functioned. Several pertinent facts have emerged from this analysis, but far more remains unexplained. First,  $\gamma/\delta$  receptors, like  $\alpha/\beta$  receptors, are stably associated on the cell surface with CD3; this has allowed use of anti-CD3 to precipitate  $\gamma/\delta$  from the cell surface as Goodman and Lefrançois report in this issue<sup>1</sup>. Such cells lack the CD4 and CD8 molecules found on virtually all CD3  $\alpha/\beta$  T cells, and are thus referred to as double negative (DN) T cells. Second, in most lymphoid tissues, cells bearing  $\gamma/\delta$  receptors are a very minor cellular component, comprising a few per cent of cells in the thymus, and less than 1 per cent in most other lymphoid tissues. It was thus a real surprise when it was discovered by the groups of Stingl<sup>11</sup> and of Tigelaar<sup>12</sup> that a dendritic, Thy-1<sup>+</sup> cell (DEC) in murine epidermis bears exclusively  $\gamma/\delta$  receptors. This result, combined with data from many investigators showing the presence of lymphocytes within most surface epithelia<sup>13</sup>, led my colleagues and I to propose that most epithelial lymphocytes would express  $\gamma/\delta$  receptors, and that this class of T cells would be involved in surveillance of epithelial cell surfaces<sup>14</sup>.

This hypothesis is given significant impetus by the findings of Goodman and Lefrançois<sup>1</sup>. These investigators extend the earlier observations on DEC to intestinal intraepithelial lymphocytes (IEL), showing that such cells have predominantly CD3-associated  $\gamma/\delta$  receptors. Unlike DEC and somatic (within the body)  $\gamma/\delta$  T cells, IEL have surface CD8 (although the cells first shown to have  $\gamma/\delta$  receptors by Brenner and colleagues<sup>4</sup> were also partly CD8<sup>+</sup>). Furthermore, S. Tonegawa and M. Bonneville reported identical findings at a recent meeting<sup>†</sup>, and also showed that such receptors use a  $V_\gamma$  and  $V_\delta$  different from that reported at the same conference by R. Tigelaar and by J. Allison to be expressed on virtually all DEC. A. Hayday and colleagues (personal communication) have obtained the same result using *in situ* hybridization. This last result is very important, as the isolation procedures used by Tonegawa and by Goodman and Lefrançois might not yield a truly representative population of IEL. Cooper and colleagues have used immunocytochemistry to show that IEL in both chickens and man are predominantly CD8<sup>+</sup> T cells with  $\gamma/\delta$  receptors. Thus there is evidence from several laboratories and at least three species that the intestinal epithelium

\*"We shall defend our island whatever the cost may be, we shall fight on the beaches, we shall fight on the landing grounds, we shall fight in the fields and in the streets, we shall fight in the hills, we shall never surrender." Speech, 4 June 1940.

†FASEB annual meeting, Las Vegas, 1–5 May 1988.

contains a resident lymphocyte population largely comprising CD8<sup>+</sup>  $\gamma/\delta$  T cells.

These findings have several important implications. First, there must be a specific mechanism by which CD8<sup>+</sup>  $\gamma/\delta$  cells using a particular  $V_\gamma/V_\delta$  receptor home to intestinal epithelium, while a second, distinct population of CD8<sup>+</sup>  $\gamma/\delta$  T cells using a different  $V_\gamma$  and  $V_\delta$  homes to the epidermis. Thus, this part of the immune system has a very precisely defined anatomy. This impression of structure is reinforced by examining the highly regular distribution of lymphocytes in epithelia, particularly of DEC in skin. Second, whatever function these  $\gamma/\delta$  T cells are serving in epithelia, they are likely to be best suited to it, as  $\alpha/\beta$  T cells are rare in these sites. Third, as previously observed for activated  $\gamma/\delta$  T cells in man, and with activated IEL, Goodman and Lefrançois<sup>1</sup> now show that IEL have a cytolytic function that is inducible in short-term assays using receptor cross-linking antibodies as activators. It is clear that we will not understand the role of the  $\gamma/\delta$  T cell until we know the nature of the ligand recognized by this receptor.

The main difficulty in the search for the ligand recognized by this new class of T-cell receptors has been that  $\gamma/\delta$  T cells have not yet been detected as contributing to the immune response to a specific antigen.

This lack of a defined ligand has led a number of investigators to examine the structure of the  $\gamma/\delta$  T-cell receptor itself, in hopes of finding clues to the nature of its ligand. Here, four features are noteworthy: first, the  $\gamma/\delta$  receptor is similar to the  $\alpha/\beta$  receptor, including association with CD3 at the cell surface. Second, the  $\gamma$ - and  $\delta$ -gene complexes contain relatively few germline  $V$ -gene segments<sup>3,6,7,14</sup>. Third, the product of a given  $V_\gamma$  gene segment tends to pair with that of a given  $V_\delta$  gene segment, thus further restricting the diversity of this receptor. Fourth, there is significant junctional diversity in the  $\gamma$ - (ref. 15) and particularly  $\delta$ - (refs 6,7)  $V$  domains generated by somatic recombination of these genes during T-cell development. This remarkable degree of junctional diversity has led Davis and Bjorkman<sup>16</sup> to propose that the  $\gamma/\delta$  receptors will be as diverse as immunoglobulin or the  $\alpha/\beta$  T-cell receptor, and therefore that the ligand for  $\gamma/\delta$  T cells will consist of or contain foreign antigen. Because the ligand for the homologous  $\alpha/\beta$  T-cell receptor is a complex of a peptide fragment of antigen bound to an MHC molecule (see table; ref. 17), Davis and Bjorkman predict that the  $\gamma/\delta$  receptor will recognize a similar ligand. This suggestion is supported by experi-

ments of Matis *et al.*<sup>18</sup> who have reported a cloned T-cell line that recognizes an as yet unidentified MHC gene product. Recently, a second possible class of ligands for the  $\gamma/\delta$  receptor has been identified by A. Krensky in man and by H. Wortis in the mouse. These investigators find  $\gamma/\delta$  T cells responding to the variable region or idiotype of surface immunoglobulin on B cells. This is consistent with the highly diverse nature of the  $\gamma/\delta$  receptor and with the absence of CD4 and CD8, as the response does not involve the MHC molecules recognized by these proteins. Such  $\gamma/\delta$  T cells could be involved in idiotype network<sup>19,20</sup> regulation of the immune response.

### Distinct ligand

Although junctional diversity is a notable feature of  $\gamma/\delta$  T cell receptors, note that the  $V_\gamma$  genes expressed on DEC<sup>12</sup> and IEL (M. Bonneville, personal communication) are those found in the thymus primarily before birth. Elliott *et al.*<sup>7</sup> show that  $\delta$ -complementary DNAs obtained at this time have strikingly less junctional diversity than is found in adult  $\gamma/\delta$  molecules. This, together with the very limited use of  $V_\gamma$  and  $V_\delta$  in DEC and IEL, suggests that the  $\gamma/\delta$  receptors of epithelial lymphocytes have a ligand distinct from,

Compartmentalization of immunological defences

| Location of pathogen | Type of immunity                           | Receptor               | Recognition                                                | Ligand                                                   | Accessory systems         | Function                              | Location                 |
|----------------------|--------------------------------------------|------------------------|------------------------------------------------------------|----------------------------------------------------------|---------------------------|---------------------------------------|--------------------------|
| Extracellular        | Humoral immunity<br>B lymphocytes          | Surface immunoglobulin | IgM                                                        | Bacterial surface                                        | Complement                | Lysis                                 | Intravascular            |
|                      |                                            |                        | IgG                                                        | Bacterial surface<br>bacteria, viruses<br>virus, toxin   | Complement<br>phagocytes  | Lysis<br>engulfment<br>neutralization | Tissue fluids            |
|                      |                                            |                        | IgE                                                        | Antigens, allergens<br>parasitic worms                   | Mast cells<br>eosinophils | Vascular changes<br>killing           | Sub-epithelial           |
|                      |                                            |                        | IgA                                                        | Virus, toxin<br>bacterial adhesins, virus                | —                         | Neutralization<br>block adherence     | Epithelial<br>surfaces   |
|                      |                                            | CD3 $\alpha/\beta$     | CD4 <sup>+</sup> T                                         | Extracellular protein<br>fragment, class II MHC          | B lymphocyte              | B-cell activation                     | Somatic                  |
|                      |                                            |                        | CD4 <sup>+</sup> T                                         | Fragment of intracellular<br>bacteria, class II MHC      | Macrophage                | Macrophage<br>activation              | Somatic                  |
| Intracellular        | Cell-mediated<br>immunity<br>T lymphocytes | CD3 $\gamma/\delta$    | CD8 <sup>+</sup> T                                         | Fragment of cellular<br>or viral protein,<br>class I MHC | —                         | Target cell<br>killing                | Somatic                  |
|                      |                                            |                        | $V_\gamma 5/V_\delta$<br>CD4 <sup>+</sup> CD8 <sup>+</sup> | ?                                                        | ?                         | ? Killing                             | Epidermis                |
|                      |                                            |                        | $V_\gamma 6/V_\delta$<br>CD8 <sup>+</sup>                  | ?                                                        | ?                         | Killing                               | Intestinal<br>epithelium |
|                      |                                            |                        | $V_\gamma/V_\delta$<br>CD4 <sup>+</sup> CD8 <sup>+</sup>   | Ig, MHC, ?                                               | ?                         | Killing, ?                            | Somatic                  |

Mechanisms of immune defence are grouped according to the location of the pathogen, the type of cell and receptor mediating specific recognition, and the body compartment protected. Humoral immunity is transferrable to naive recipients with serum antibody, produced by B lymphocytes, while cell-mediated immune responses are transferrable only with immune T lymphocytes. The main function of humoral immunity is the destruction or inactivation of extracellular forms of microorganisms; however, within the extracellular spaces, humoral immunity is distinctly compartmentalized as shown. The different isotypes of immunoglobulin provide protection of distinct body spaces. IgM, the first antibody produced, is essentially confined to the circulation, while IgG antibodies enter tissue spaces as well. IgE antibodies bind to high-affinity receptors on the surface of the mast cell, most of which occupy a sub-epithelial location. When antigen crosses an epithelial layer and binds to this IgE, the mast cell is activated resulting in profound alterations in local circulation allowing the entry of immune reactants into the site of antigen deposition, and the transfer of antigen to lymph nodes, where immune responses are generated. Finally, IgA is actively transported across many epithelia, providing protection outside the body by blocking bacterial and viral adherence and toxin activity. Thus, the humoral immune system has evolved distinct molecules to provide protection of distinct extracellular compartments of the body. T lymphocytes have only a surface form of the receptor. This receptor is made up of two variable chains, called  $\alpha$  and  $\beta$ , disulphide linked to one another and also stably associated in the membrane with the invariant molecules of the CD3 complex. Such T cells recognize antigen in the form of peptides bound into the cleft of a molecule encoded in the major histocompatibility complex (MHC)<sup>17</sup>. Two classes of T cells, distinguished by surface expression of the molecules CD4 and CD8, recognize antigen bound by class II and class I MHC molecules, respectively<sup>18</sup>. Class II MHC molecules are found mainly on B cells and macrophages, and they are specialized to present antigens derived from outside the cell. CD4 T cells are primarily involved in activating B cells to secrete antibody in response to extracellular pathogens, and activating infected macrophages to destroy the proliferating bacteria they harbour; these may be functions of discrete CD4 T cell subsets<sup>21</sup>. Class I MHC molecules are found on most cells in the body, and are specialized to present antigens synthesized by the cell's own biosynthetic machinery. The primary role of CD8 T cells is to monitor the genetic integrity of cells, and to kill those cells infected with virus by recognizing viral peptides presented by class I MHC molecules on the cell surface. Finally, T cells bearing  $\gamma/\delta$  receptors may also mediate a variety of functions of cellular immunity, in ways discussed in detail in the text.



and much less diverse than, those recognized by  $\gamma/\delta$  receptors found in somatic lymphoid tissues. In considering potential ligands for those less diverse  $\gamma/\delta$  receptors found on DEC and IEL, anatomy may again stimulate our understanding.

The  $\alpha/\beta$  T cells that are the dominant effector cells in the body are characterized by their remarkable recirculation. These cells traverse from blood to lymphoid tissues to blood, contacting the surface of many antigen presenting cells (APC) in the lymph nodes and spleen. A pertinent question is how many T cells contact a given APC per unit time, as immune responses can occur only when a T cell encounters an APC bearing the antigen for which its receptor is specific. This number would have to be large to accommodate the low frequency of specific T cells in this recirculating pool; this is supported by the observation that all antigen-specific T cells cease recirculation within 48 hours of antigen injection. In epithelia, by contrast, the number of IEL an average epithelial cell encounters per day is probably one. If the function of IEL is to recognize alterations in epithelial cell surfaces that result from infection or transformation, then it is unlikely that IEL will be effective in this task if they recognize a set of ligands as diverse as that recognized by  $\alpha/\beta$  T cells, as the limited number of cell interactions possible in this setting would preclude recognition in all but the rarest cases.

As noted previously, the  $\alpha/\beta$  receptors are specialized for MHC recognition, so it seems likely that the  $\gamma/\delta$  receptor will be as well. However, there is no obvious reason to suppose that  $\gamma/\delta$  receptors will recognize products of the same MHC genes as do  $\alpha/\beta$  receptors. The expression of CD8 on  $\gamma/\delta$  IEL<sup>1</sup> further strongly suggests that these cells will be specific for class I MHC molecules, as CD8 expression is associated with class I MHC recognition (table 1; ref. 21), and isolated CD8 binds class I MHC molecules. The bulk of  $\alpha/\beta$  T cells so far characterized recognize class I and class II MHC molecules. It seems possible that  $\gamma/\delta$  T-cell receptors will be specific for the many uncharacterized class I-like genes that map distal to *H-2D* in the mouse, called class IB by Klein<sup>22</sup>. This supposition is supported by results obtained using two different cloned  $\gamma/\delta$  T cells that recognize class IB MHC molecules (ref. 18; S. Tonegawa); however, it should be pointed out that J. Bluestone, L. Matis and colleagues have also cloned  $\gamma/\delta$  T cell lines that are specific for class I and class II MHC molecules when allogeneic spleen cells are used as stimulators. I contend that this stimulus, although effective, is eliciting rare cross-reactions that may not be representative of the self molecules that are likely to be a critical part of the ligand for  $\gamma/\delta$  T cell receptors.

The  $\gamma/\delta$  T-cell recognition of autologous

class IB MHC molecules expressed by epithelial cells under conditions of stress might be an effective if seemingly primitive means of epithelial defence. These genes, unlike class I MHC genes, are expressed in a tissue- and activation-specific fashion in those cases in which this has been examined. As shown by Goodman and Lefrançois<sup>1</sup>, this recognition event should lead to killing of the epithelial cell. This form of defence at the immunological frontiers, outside the epithelial basement membranes, could be highly effective. Epithelia that lose a cell or two rapidly cover such lesions by lateral migration of healthy cells into the open area. As long as an infected or a transformed cell is killed before spread across the basement membrane can occur, neither infection nor malignancy can ensue. This system may lack the sophistication of antigen specificity and immunological memory, but it may also not need them to perform its function. It is tempting to speculate that this present-day frontier is also the site in which cell-mediated immunity originated, and that this 'new' component of the immune system is actually one of its oldest. The immune system has engrafted antigen specificity onto several existing non-specific effector mechanisms, and the  $\gamma/\delta$  T-cell system may exemplify this principle again.

It will be interesting to learn more about epithelial lymphocytes, the heterogeneity of their receptors, their functions and their ligands. From this, we may be able to deduce their overall biological role. Deficiencies in these cells, if discovered, could also be enlightening as to the importance of this outermost barrier of cell-mediated immunity. It seems likely that it will be the  $\gamma/\delta$  T cell that will stand and be counted when it comes to the defence of our epithelial frontiers. □

1. Goodman, T. & Lefrançois, L. *Nature* **333**, 855-858 (1988).
2. Saito, H. *et al.* *Nature* **309**, 757-762 (1984).
3. Hayday, A.C. *et al.* *Cell* **40**, 259-269 (1985).
4. Brenner, M.B. *et al.* *Nature* **322**, 145-149 (1986).
5. Bank, L. *et al.* *Nature* **322**, 179-184 (1986).
6. Chien, Y.-H. *et al.* *Nature* **327**, 677-682 (1987).
7. Elliott, J.F. *et al.* *Nature* **331**, 627-631 (1988).
8. Hata, S., Brenner, M.B. & Krangel, M.S. *Science* **238**, 678-682 (1987).
9. Band, H. *et al.* *Science* **238**, 682-684 (1987).
10. Loh, E.Y. *et al.* *Nature* **330**, 569-572 (1987).
11. Koning, F. *et al.* *Science* **236**, 834-836 (1987).
12. Kuziel, W.A. *et al.* *Nature* **328**, 263-266 (1987).
13. Ernst, P.B., Befus, A.D. & Bienenstock, J. *Immun. Today* **6**, 50-55 (1985).
14. Janeway, C.A. Jr, Jones, B. & Hayday, A.C. *Immun. Today* **9**, 73-76 (1988).
15. Jones, B., Mjoiness, S., Janeway, C.A. Jr & Hayday, A.C. *Nature* **323**, 635-637 (1986).
16. Davis, M.M. & Bjorkman, P.J. *Nature* (in the press).
17. Bjorkman, P.J. *et al.* *Nature* **329**, 512-517 (1987).
18. Matis, L.A. *Nature* **330**, 262-264 (1987).
19. Jerne, N.K. *Ann. Immun.* **125C**, 373-389 (1974).
20. Bottomly, K. *Immun. Rev.* **79**, 45-61 (1984).
21. Swain, S.L. *Immun. Rev.* **74**, 129-156 (1983).
22. Klein, J. *Natural History of the Major Histocompatibility Complex* (Wiley, New York, 1986).
23. Janeway, C.A. Jr *et al.* *Immun. Rev.* **101**, 39-80 (1986).

Charles A. Janeway Jr is in the Section of Immunology and Department of Pathology and Biology, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06510, USA.

## Daedalus

# Dreams for sale

THE brain differs from a computer in at least two important ways. It seems not to store data in fixed locations; and it hangs onto that data tenaciously (whereas we all know how disastrously easy it is to erase computer-stored data by overwriting it). Daedalus muses that the brain must store its data dynamically, like a magnetic-bubble memory or a charge-coupled device. Entering new data does not overwrite the old stuff, but pushes it along a bit to make room.

This notion ties in neatly with the claim that a dream is the repacking of the day's experiences: the updating of the current world-picture. The discarded data are not overwritten, but pushed out of the way. In fact they are pushed right back into the sensory region of the brain, to be sensed as a dream.

And this, says Daedalus, is why dreams (to the dismay of psychiatrists) are usually such nonsense. They are the rejects on the brain's cutting-room floor, the discards of the night's editing. Furthermore, they are in the brain's internal machine-code, and like all machine-codes can be interpreted as high-level language only with the aid of many arbitrary symbols.

But Daedalus takes his theory further. He recalls the characteristic rapid eye movements of dreaming, as if behind closed eyelids the dreamer were scanning a visual image. Maybe the rejected dream images are pushed, not just back into the sensory processing area, but right down the optic nerves to the retinas, where the eyes' reflex motor responses try to follow them.

To test this appealing notion, Daedalus is devising special telemetric retinal-imaging goggles for DREADCO volunteers to wear in bed. The goggles can be filled with simulated tear fluid. With good fortune this should not prevent the bathed eyes from closing in sleep, but will complete an ultrasonically clear path from each retina out through eye, eyelid and fluid to a focusing lens and piezoelectric detector array in the goggles. The idea is that, because a nerve expands slightly on transmitting an impulse, the dreaming retina should radiate an ultrasonic pattern corresponding to the image it is 'seeing'. The piezoelectric goggles will intercept that dream image, and transmit it for display on a television monitor.

Thus the poignant terrors, charms and crazinesses of our nocturnal fantasies will be captured for objective study. Brain theorists, psychiatrists and dreamers themselves will be fascinated. Film and television companies will rush to screen the wilder dreams: their surrealistic visual vocabulary will speak compellingly to our subconscious minds.

David Jones

## Advantages of heat-stable vaccines

SIR—I agree with Arya<sup>1</sup> that a desired criterion for new chimaeric poliovirus vaccines should be stability to warm temperatures. It should be pointed out, however, that live poliovirus vaccines that contain a thermal stabilizer have been available from several manufacturers and in use in many countries, including the United Kingdom, for more than 20 years.

In the early 1960s it was found that the infectivity of the enteroviruses could be preserved even when they were heated at 50 °C, if molar MgCl<sub>2</sub> was added<sup>2</sup>, a property that is still used in their identification and characterization. This finding was quickly applied to the live vaccine<sup>3</sup>, not only in laboratory manipulations<sup>4</sup> but also in the field, where stabilized vaccines, in the absence of a cold chain, were used effectively to halt type 1 and type 3 outbreaks in a semitropical country<sup>5</sup>.

In mass poliovirus immunization programmes, particularly in developing tropical countries where it is difficult to maintain or to transport the vaccine under frozen conditions, the stabilized vaccine offers numerous advantages<sup>6</sup>.

In laboratory studies with live poliovaccine stored in MgCl<sub>2</sub>, investigators at Wellcome Research Laboratories<sup>7</sup> found that the vaccine showed so little loss in virus titre after long-term storage at -20 °C that the predicted half-life was calculated as 92 years. They further found that MgCl<sub>2</sub>-stabilized vaccine suffered no significant loss of potency after as many as nine cycles of alternate warm and cold conditions. The samples were alternately exposed to ambient temperatures and to storage frozen at -20 °C or liquid at 7-9 °C.

Although this handling certainly cannot be recommended as a general practice, it is reassuring to know that properly stabilized vaccine can survive some of the 'cold chain' problems that have arisen in field use. Similar results obtained by another manufacturer<sup>8</sup> indicate that MgCl<sub>2</sub>-stabilized vaccine retains potency in field and shipping conditions without refrigeration for 2 weeks, including periods when the daytime temperature reached 42 °C.

At a recent meeting<sup>9</sup>, it was pointed out that consistent use of stabilizers such as MgCl<sub>2</sub> would help minimize reliance on the cold chain. Such stabilized vaccine, although not perfect, would meet the suggested target of stability for 3 days at 37 °C with a maximum loss in potency of 0.5 log.

JOSEPH L. MELNICK

Department of Virology and  
Epidemiology,  
Baylor College of Medicine,  
Houston,  
Texas 77030,  
USA

## Myosin heads in muscle thick filament assembly

SIR—Morel and Bachouchi<sup>1</sup> discuss the role of myosin heads in the assembly of skeletal muscle thick filaments. They cite evidence to show that the binding of MgATP to myosin heads is crucial in the assembly of thick filaments of physiological diameter. Although data in support of these head-mediated phenomena appear reasonable, they may be irrelevant to the generation of native-like filaments from myosin.

Synthetic myosin filaments assume different morphological forms depending on the nature of the filamentogenic solvent. The experiments cited by Morel and Bachouchi were carried out on what are conventionally called the pH 7 class of filaments. These differ from *in vivo* filaments in that they are generally spindle-shaped, thicker and lack a clearly demarcated bare zone at their centres<sup>2,3</sup>.

As pointed out<sup>1</sup>, the addition of MgATP to the pH 7 class of filaments does appear to make them more native-like in structure. On the other hand, this treatment seems to be unnecessary with filaments of the pH 8 class. These filaments are morphologically similar to native filaments save that they are shorter and have a narrow length-distribution rather than the precise length seen *in vivo*<sup>2,3</sup>. Interactions between the myosin heads and the filament core seem to be minimal, as the heads and part of the tail (the S-2 segment) are predominantly dissociated free of the filament surfaces<sup>4,5</sup>. This observation effectively excludes head-core interactions from playing a significant role in their assembly.

The limit on length can be overcome by copolymerizing the myosin with C-protein<sup>6</sup>, a constituent of the native filament. The longer filaments so formed retain the physiological diameters typical of pH 8 and native filaments. The only gross structural property to distinguish them from native thick filaments is the lack of a precise length. A crucial role for head-head and head-core interactions

and/or for MgADP or MgATP in assembly therefore seems unlikely. It is more plausible that the head-related phenomena function in contraction than in assembly (see ref. 7).

JULIEN S. DAVIS

Department of Biology,  
The Johns Hopkins University,  
34th and Charles Streets,  
Baltimore,  
Maryland 21218, USA

1. Morel, J. E. & Bachouchi, N. *Nature* **332**, 591 (1988).
2. Kaminer, B. & Bell, A. L. *J. molec. Biol.* **20**, 391-401 (1966).
3. Josephs, R. & Harrington, W. F. *Biochemistry* **5**, 3474-3487 (1966).
4. Sutoh, K., Chiao, Y.-C. C. & Harrington, W. F. *Biochemistry* **17**, 1234-1239 (1978).
5. Reisler, E. & Liu, J. *J. molec. Biol.* **157**, 659-669 (1982).
6. Davis, J. S. *J. Muscle Res. Cell Motil.* (in the press).
7. Davis, J. S. *A. Rev. Biophys. biophys. Chem.* **17**, 217-239 (1988).

## Designing antibodies for human therapies

SIR—The recent descriptions<sup>1,2</sup> of reshaping of human antibodies represent remarkable feats of genetic engineering with potential to produce reagents useful in human therapy. It is anticipated that these would be superior to rodent antibodies as immunotherapeutic agents because treated individuals should not develop antibodies to the human variable-region framework and constant regions of the heavy (H) and light (L) chains of immunoglobulin molecules. The concern remains, as explicitly expressed by Riechmann *et al.*<sup>1</sup>, that the idiotypic portion of the reshaped human IgG1 antibody that they hope to use in therapy may be immunogenic.

In that regard, I wish to raise the issue of H- and L-chain Gm and Km allotypes. Although antibody responses to the Km and Gm allotypes may not themselves present major problems, the presence of foreign allotypes may provide carrier determinants and thus enhance immune responses to idiotypes. In inbred strains of mice, immunization with myeloma proteins bearing allotypes foreign to the recipient strain elicits anti-idiotypic more readily than when syngenic strains are immunized<sup>3</sup>.

I therefore suggest that those planning clinical studies consider Km and Gm matching their recipients with the therapeutic antibodies. Both the reported human IgG1 clones<sup>3,4</sup> have sequences corresponding to the Gm1 type in CH3 (Arg-Asp-Glu-Leu at 355-358) and to Gm17 (Lys 214) in CH1. Because of the success of site-directed mutagenesis experiments<sup>1,2</sup> it should be simple to introduce the necessary point mutations to produce reagents with the alternative allotypic forms of human IgG1 and  $\kappa$ -light chains. It will be considerable interest and benefit if allotype-matching prolongs the time period of effective

1. Arya, S. C. *Nature* **332**, 789 (1988).
2. Wallis, C. & Melnick, J. L. *Virology* **16**, 504-505 (1962).
3. Melnick, J. L., Ashkenazi, A., Midulla, V. C., Wallis, C. & Bernstein, A. *J. Am. med. Ass.* **185**, 406-408 (1963).
4. Petersen, I. & von Magnus, H. *Acta Path. Microbiol. Scand.* **61**, 652-653 (1964).
5. Yofe, J., Goldblum, N., Eylan, E. & Melnick, J. L. *Am. J. Hyg.* **76**, 225-238 (1962).
6. Mirchamsy, H., Shafiyi, A., Mahinpour, M. & Nazari, P. *Vaccinations in Developing Countries: Dev. Biol. Standardiz.* **41**, 255-257 (Karger, Basel, 1978).
7. Finter, N. B., Ferris, R., Kelly, A. & Prydie, J. *Vaccinations in Developing Countries: Dev. Biol. Standardiz.* **41**, 271-276 (Karger, Basel, 1978).
8. Peetermans, J., Colinet, G. & Stephenne, J. *Proc. Symp. Stability and Effectiveness of Measles, Poliomyelitis, and Pertussis Vaccines*, 61-66 (Yugoslav Acad. Sci. Arts, Zagreb, 1978).
9. Institute of Medicine, National Academy of Sciences *Temperature Stable Vaccines for Developing Countries: Significance and Development of Strategies* (Washington, DC, 1987).



therapy by diminishing anti-idiotypic responses.

ROSE G MAGE

Laboratory of Immunology,  
National Institute of Allergy and Infectious  
Diseases,  
National Institutes of Health,  
Bethesda, Maryland 20892, USA

1. Riechmann, L. *et al.* *Nature* **332**, 323–327 (1988).
2. Verhoeven, M., Milstein, C. & Winter, G. *Science* **329**, 1534–1536 (1988).
3. Ellison, J.W., Berson, B.J. & Hood, L.E. *Nucleic Acids Res.* **10**, 4071–4079 (1982).
4. Takahashi, N. *et al.* *Cell* **29**, 671–679 (1982).
5. Mage, R.C. *et al.* in *The Antigens* (ed. Sela, M.) 341 (Academic, New York, 1973).

## The nature of Taung dental maturation continued

SIR—Mann questions<sup>1</sup> the emerging consensus from diverse lines of evidence that australopithecines had more ape-like dental development patterns, and thus more rapid maturation rates, than modern humans<sup>2,3</sup>. He now<sup>4</sup> specifically questions our interpretation of the developing dentition of the Taung skull as ape-like (as deduced from high-resolution CT scans<sup>5</sup>), proposing instead that some modern *Homo sapiens* have a dental pattern similar to Taung. He thus denies that apes and humans can be distinguished in terms of such dental patterns.

By way of evidence he illustrates 1 of 25 immature dentitions from the 3,000-year-old archaeological site at Hasanlu, Iran. As we assume he has chosen the best possible specimen in support of his view, it is surprising that it provides further support to our interpretation of the Taung skull. Furthermore, this specimen lacks many of the critical features needed to address the similarity of contemporary human dentitions with Taung: (1) with the exception of the left M1, not a single

permanent tooth remains of the jaw; (2) because they are all isolated teeth, their position relative to one another and to the alveolar margin is impossible to judge; and (3) the lower central incisors of this specimen may have already erupted, or these teeth may have come from more than one individual.

Even disregarding these problems, we find little credibility in the claim that this specimen mimics dental development in Taung. Mann himself recognizes this<sup>4</sup> in the last two sentences of the legend to his Fig. 2. The differences he notes in upper and lower incisor-root formation and in second-molar calcification completion between the Iranian specimen and Taung are exactly those expected and predicted between ape and human-like dental patterns at this particular M1 developmental stage.

Our figures verify this point. Figure 1 shows 2-mm parasagittal CT scans of a chimpanzee, a human child, and the Taung skull, all at the same M1 dental maturation stage. Note that the permanent incisors of the chimpanzee and Taung skull are virtually identical in their horizontal alignment, their distance from the incisal alveolar margin and their total lack of root development (lower incisors also lack root development). By contrast, the permanent central incisor in the human child is vertically aligned, at or near the alveolar margin, and has a well-developed root. Figure 2 also shows clearly that the ape and Taung are virtually identical in the orientation and degree of calcification of the second molar—neither approaches the complete crown calcification of the M2 in the Iranian specimen described by Mann<sup>4</sup>.

Certainly, no biological anthropologist can seriously argue that dental develop-

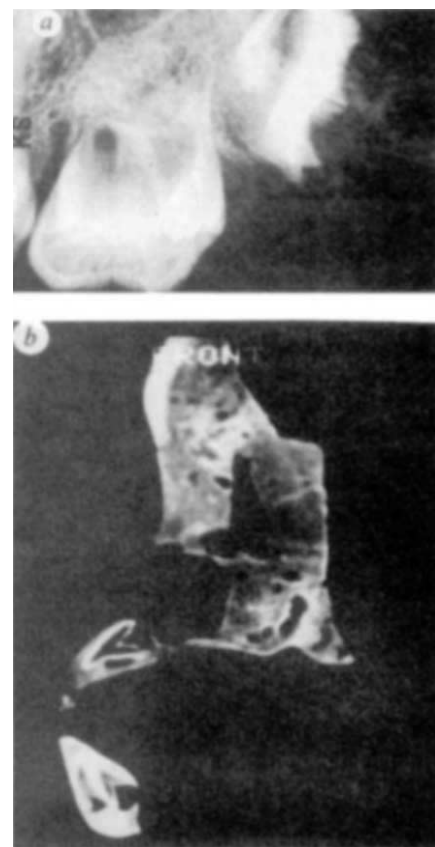


Fig. 2 a, An X-ray of the M1 and developing M2 in an orangutan at the same M1 development stage as the Taung skull. b, A parasagittal CT scan through the developing M2 in Taung.

ment in apes and humans proceeds along the same trajectory. That is not to say, however, that some dental stages may be equivocal in distinguishing between the two, but this is certainly not the case with the stage of first molar eruption in the Taung skull<sup>6,7</sup>.

Mann<sup>4</sup> concludes from his study of the Iranian sample: either “many modern humans grow in an ape-like pattern; or the patterns proposed and used . . . are incapable of distinguishing apes from humans and cannot be applied to characterize the nature of development in fossil hominid specimens”. We reject the notion that: (1) modern Iranians mimic apes in their maturation patterns; or (2) that dental development patterns are incapable of distinguishing humans from apes.

GLENN C. CONROY

Department of Anatomy & Neurobiology  
and Department of Anthropology

MICHAEL W. VANNIER

Mallinckrodt Institute of Radiology,  
Washington University Medical Center,  
St Louis, Missouri, 63110, USA

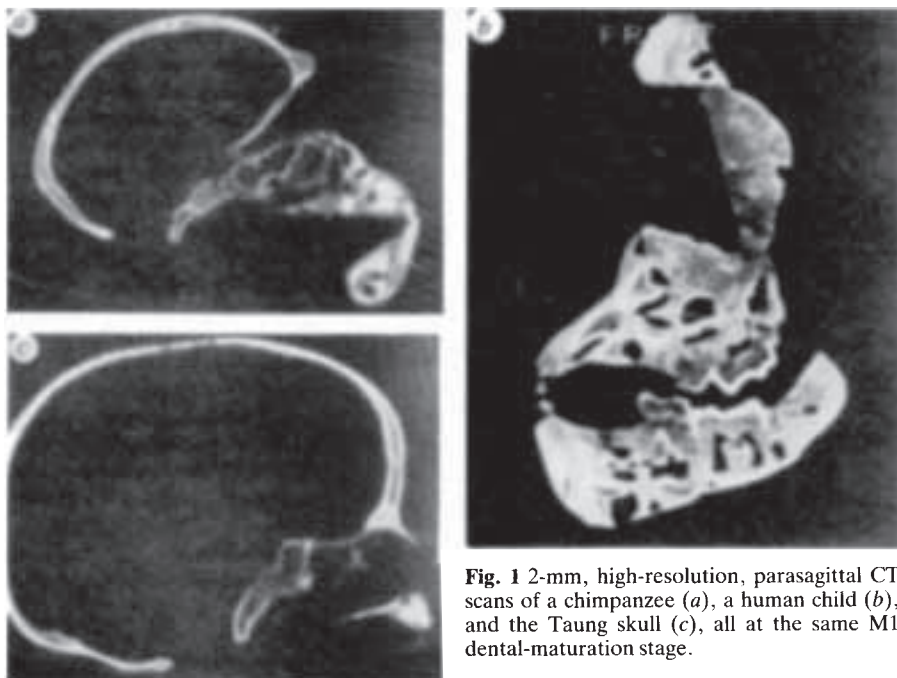


Fig. 1 2-mm, high-resolution, parasagittal CT scans of a chimpanzee (a), a human child (b), and the Taung skull (c), all at the same M1 dental-maturation stage.

1. Mann, A.E. Lampl, M. & Monge, J. *Nature* **328**, 673–674 (1987).
2. Bromage, T.G. & Dean, M.C. *Nature* **317**, 525–527 (1985).
3. Smith, B.H. *Nature* **323**, 327–330 (1986).
4. Mann, A.E. *Nature* **333**, 123 (1988).
5. Conroy, G.C. & Vannier, M.W. *Nature* **329**, 625–627 (1987).
6. Dean, M.C. & Wood, B.A. *Folia Primatol.* **36**, 111–127 (1981).
7. Bromage, T.G. *J. Hum. Evol.* **16**, 257–272 (1987).

# Sir Joseph in the round

David E. Allen

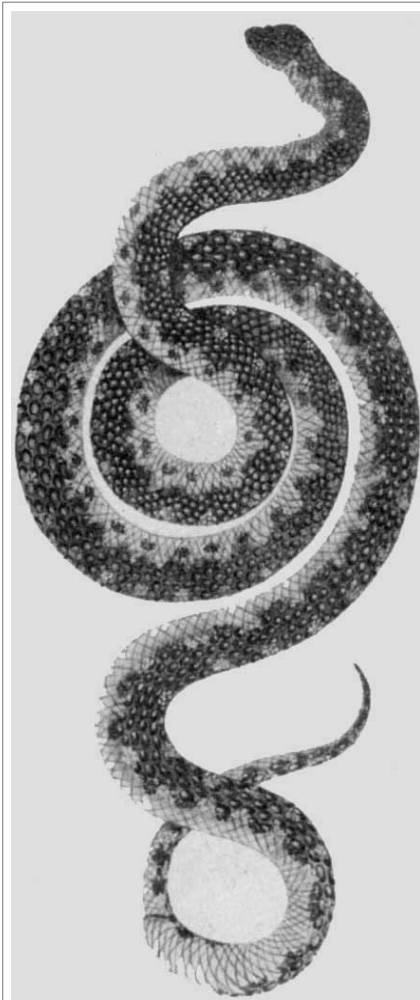
**Sir Joseph Banks 1743–1820.** By Harold B. Carter. *British Museum (Natural History)\**: 1988. Pp.671. £45.

VICTORIANISM is back in very truth! Here is a huge slab of biographical masonry of just the kind in which our forebears revelled, complete with the standard introductory chapter detailing the subject's ancestry, 27 appendices, an extensive bibliography and an index that runs to 54 pages.

It comes as no surprise to learn from the foreword that more than a quarter of a century has gone into the book's creation and into the massive archival enterprise that underlies it. It is the latter indeed that, down the years, has been the author's main concern. Catastrophically for later research, Banks's papers were dispersed at his death and until now all accounts of this pivotal figure of late eighteenth-century science have been based on fragmentary evidence. Mr Carter's signal achievement has been to bring together in one place — namely, the British Museum (Natural History), the primary recipient of the Banksian *reliquiae* — a collection of 15,000 documentary items, either originals or copies, tracked down from all around the globe. No one would have blamed him if he had seen his task as ending with that. But fortunately he has been moved to be more than an archivist and has produced in addition a complementary quarry for researchers that is as impressive in its mastery of detail as it is well written.

Banks was President of the Royal Society, without a break, for three-quarters of his adult life — 42 years in all. That is an extraordinary achievement in itself. But even without the advantages of that office, his farsightedness and his determination allied to his personal wealth would have been sufficient to make his life a success-story in any age. As it was, he had the extra measure of good fortune to be born into a period in which the particular branch of knowledge that took his interest was about to offer its greatest-ever scope to those prepared to pursue it. The second half of the eighteenth century was to be the great age of botany, the great age of private patronage and the great age of European exploration. Banks seized all three opportunities with aplomb.

The wealth came from landed estates, principally in Lincolnshire, the product of some shrewd investments by an ancestor of characteristic Banksian girth. Banks succeeded to these at the early age of 18, by which time he had already begun to manifest a broad interest in natural history. His years at university seem to have been in the main an irrelevance and the key event, when he was 21 or thereabouts,



*Twist in the tale — this diamond python was painted by an artist who landed in New South Wales with the First Fleet in 1788. The works of this unidentified artist, sometimes called the 'Port Jackson Painter', were part of the library of Sir Joseph Banks, and are known as the Banks Ms34. The picture is taken from a colour original in The Art of the First Fleet and Other Early Australian Drawings, a beautifully presented book edited by B. Smith and A. Wheeler and published by Yale University Press, price \$125, £95 (£79.95 until September).*

was probably his introduction to Daniel Solander, the British Museum official and former pupil of Linnaeus who was to become his lasting friend, employee and co-worker.

Doors opened readily, and election as a Fellow of the Society of Antiquaries and then of the Royal Society occurred as a matter of course. But set on being much more than the mere *dilettante* that so many other young men of his bent and background were content to be, Banks proceeded to put himself through a rigorous five-year apprenticeship of scientific travel. The centrepiece was the epoch-making voyage of 1768–1771 to the South Seas in the *Endeavour* under Cook, with its haul of over 30,000 specimens (among them perhaps over 1,400 plant species new to science); but in addition there were journeys to Newfoundland and Labrador and to Iceland, of ten and five months respectively, on either side. Four other expeditions to distant parts got no further than the planning stage.

By the age of 30 all of Banks's travelling was done, and he settled down to the working up of his vast collections and to launching scientific initiatives, of a steadily growing variety, from his Presidential chair. Election to the Council of the Royal Society had come two years after his return in the *Endeavour*, and scarcely five years later he was firmly at its head. Even before that he had successfully persuaded George III to start financing that long series of collector-explorers whose introductions were so greatly to enrich the Royal Gardens at Kew (and Banks's personal collections as well).

His appetite whetted, he was soon off on a programme of vicarious expeditions to almost every part of the fast-expanding British Empire, organizing the despatch of breadfruit to the West Indies, of Merino sheep to New South Wales, of tea plants and cochineal insects to India. In 1783, a particularly momentous year, we find him simultaneously exploring the possibility of a white settlement in the Pacific and capturing the Linnaean collections for Britain through the surrogacy of J.E. Smith (his own finances having taken a temporary tumble from the slump in the wool trade brought on by the American War). By 1790 he was operating on a wider stage still, serving on the Board of Agriculture, dominating the Board of Longitude and, in a bizarre interlude, becoming the chosen stalking-horse of a group of Hungarian dissidents anxious to enlist the British government in its cause.

And so it went on. Despite martyrdom to gout, the initiatives were unceasing, the intercessions ever bolder and more various. There has probably never been a scientific administrator-cum-politician so widely influential and so practically effective — nor a scientist who has won

\*In the United States both books (see over) are available from ISBS, 5602 NE Hassals Street, Portland, Oregon 97213. *Sir Joseph Banks Bibliography* is published by St Paul's Bibliographies, 1 Step Terrace, Winchester, Hampshire, price £45.



such lasting renown with scarcely a single publication to his name.

Mr Carter does rather overdo the detail (did we really need to have in print a list of the species caught by Banks's fishing parties?) and the decision to omit all references was an unhappy one. I remain unpersuaded by the author's excuse that the nature of the text makes these superfluous. Although we are assured that the source material is detailed in a companion volume, that comes from a different stable and is enough of a separate enterprise not to have been sent out for review concurrently (see footnote on previous page). Inclusion of that extra matter would no

doubt have made for a greater bulk than could be contained within a single pair of covers, but the requirements of scholarship ought not to be overridden by conveniences of production, particularly in the case of a volume that appears under such august sponsorship.

That defect is a serious one. But, all in all, here is a book that henceforward will put the work of Banksian specialists on an altogether different footing. □

*David E. Allen is in the Unit of History of Medicine, University College London, Gower Street, London WC1E 6BT, UK, and author of The Naturalist in Britain: a Social History (Allen Lane, 1976).*

## Gloom – or boom?

Norman Myers

**The Cassandra Conference: Resources and the Human Predicament.** Edited by Paul R. Ehrlich and John P. Holdren. Texas A & M University Press, Drawer C, College Station, Texas 77843: 1988. Pp.330. Hbk \$32.50; pbk \$14.95.

CASSANDRA, daughter of King Priam of Troy, was empowered with the capacity of prophecy by the god Apollo, provided she slept with him. Cassandra didn't deliver, and Apollo imposed a curse on her: her prophecies, though accurate, would be spurned as beyond belief. Hence her warnings about the Trojan horse went unheeded.

So the title of this book is not so perverse as it might seem, despite the modern connotation of Cassandra as a mere doomsayer. The contents seek to demonstrate how far we are indeed paying a price for ignoring warnings about threats to the environment. It further shows how we can still change course towards ecological accord with the biosphere, generating all manner of positive payoffs. So persuasive are the prescriptions for a rehabilitated future, that the book could well have been subtitled *Environmental Gloom – Or Boom?*

The organizers of the Cassandra conference brought together a number of leading American analysts who, for a decade or more, have been telling of a series of interrelated threats to our biosphere. Especially useful is the upgrading of earlier arguments, deploying the latest data. The thematic treatments are rigorous throughout, and there is plenty of discussion of opposing opinions, notably those of Julian Simon and the late Herman Kahn.

Garrett Hardin opens with an assessment of the population predicament. Steve Schneider looks at climate change and its links with food production. David Pimentel appraises industrial agriculture,

and related topics such as topsoil and water stocks, while Anne Ehrlich examines agriculture in the Third World. Dan Luten deals with energy, Peter Raven with tropical rainforests, John Harte with acid rain and Cheryl Holdren with toxic substances. George Woodwell considers biotic services and the human estate. Ehrlich himself rehearses the putative phenomenon of nuclear winter, while Holdren presents some innovative and hard-nosed strategies to avoid nuclear war.

This list might evoke a reaction of "Here we go again. . . . What's new?". There is lots new: the book is far from a predictable account of what is going amiss. On the population front, for example, we encounter not only the usual dismal figures (and let's note that when we discount the remarkable achievement of China, the growth rate in the Third World is still 2.4 per cent per year, with no decline recorded for several years). We also find there is much scope to cut back on fertility rates, as witness the accomplishments of South Korea, Taiwan, Java, Thailand, Kerala State in India, Cuba, and urban communities in Colombia and Mexico — countries with a wide range of economic performance, political systems and cultures. The United Nations median projection postulates establishment of 'replacement fertility' by the year 2025, leading to an ultimate total world population of well over 10 billion people. Were all developing countries to match the fertility declines of the countries cited, the two-child family could be attained by the year 2005 (a formidable but not impossible challenge), reducing the eventual total to 8 billion. Were many more countries to lose ground, as is currently the case with the Philippines, several Muslim countries and most of black Africa, the date would be deferred till 2045, with an eventual total of well over 12 billion. Thus a gap of 40 years would make a difference of more than 4 billion people.

There is similar potential for breakthroughs in the energy field. Since 1979 and the second oil shock, which was

followed by an outburst of energy conservation, the United States has expanded its economy by more than 30 per cent while consuming less energy overall. Many more energy-saving technologies remain to be mobilized. The most abundantly available response to energy shortages is that which is cheapest, most widely available and least exploited — simple saving of energy. Yet following the third oil shock of recently reduced prices, the energy-conserving momentum of the early 1980s is being lost.

All of the book's contributors devote a good part of their chapters to a prescriptive analysis of how we can reduce or avert the more severe environmental problems. For instance, Schneider sub-titles his chapter "Signs of Hope, Despair, and Opportunity". In addition there is a chapter by a World Bank economist, Herman Daly, on the steady-state economy (not necessarily the same as the no-growth economy), and how it could be established.

In an epilogue, Ehrlich and Holdren emphasize that just as the problems are interactive, so our responses should be interrelated. A solution to one problem will work only if parallel solutions are addressing many, if not most (and sometimes all), of the other problems as well. The most salient instance is overpopulation, which both contributes to poverty and reflects it. Just as human poverty and environmental impoverishment compound one another, so together they compound the chances of international conflict and nuclear war; and financial resources spent on armaments and other military dispositions are no longer available to restore forests, rehabilitate watersheds, replenish soil fertility, reduce pollution, produce more food and so forth. Which of the funding options will purchase more real and enduring security? □

*Norman Myers, Upper Meadow, Old Road, Headington, Oxford OX3 8SZ, UK, is a consultant on development and environmental matters.*

### New in paperback

- *The Extraterrestrial Life Debate 1750–1900: The Idea of a Plurality of Worlds from Kant to Lowell* by Michael J. Crowe. Publisher is Cambridge University Press, price is £15, \$22.95. For review see *Nature* **325**, 117 (1987).
- *Odd Perceptions* by Richard L. Gregory. Publisher is Routledge, price is £8.95. For review see *Nature* **325**, 206 (1987).
- *Origins: A Skeptic's Guide to the Creation of Life on Earth* by Robert Shapiro. Publisher is Penguin, price is £5.95. For review see *Nature* **320**, 646 (1986).
- *The Discovery of Insulin* by Michael Bliss. Publisher is Faber/Chicago University Press, price is £6.95, \$10.95. For review see *Nature* **303**, 261 (1983).
- *Bird of Passage: Recollections of a Physicist* by Rudolf Peierls. Publisher is Princeton University Press, price is \$12.50, £7. For review see *Nature* **320**, 660 (1986).

# Stopping short of history

John Dunster

**Chernobyl: The Real Story.** By Richard F. Mould. Pergamon: 1988. Pp. 256. Hbk £25, \$45; pbk £9.95 \$17.95.

AN AUTHOR who chooses, or lets his publisher choose, the subtitle "the real story" sets himself a high standard of historical insight, especially when the story is necessarily filtered through the information system of the Soviet Union.

Some of the information released by the Soviet authorities about the accident at Chernobyl can be checked against environmental findings in Western Europe, but there is no real information about the events on site beyond that officially released. The large, and predominantly cynical, group of scientists and engineers who listened to the Soviet presentation in Vienna in late August, 1986, were impressed both by the Soviet delegation, led by Adademician Legasov, now regrettably dead, and by the amount of written detail that had been assembled in only four months. I doubt if we in the West could have done as well in that time. Nevertheless, although the information is consistent with the basic facts of reactor physics and engineering, we must remember that it is not subject to effective challenge. Much was left unsaid and the account given of some of the events sounded suspiciously tidy.

That is the background against which this book has to be judged. The key questions must be — Does it illuminate the Soviet story? Does it probe the weaknesses or ambiguities? And, finally, does it identify the lessons for the future? Unfortunately, the answers to all these questions seem to be 'no'.

So what did happen at Chernobyl and what does Richard Mould tell us about it? The accident started with a simple and useful experiment aimed at finding out how effective the kinetic energy of the turbo-alternators would be at providing a short period of emergency power if the station lost all internal and external sources of electricity. The experiment called for the reactor to be run at about half its rated power and for the steam valve to the turbine then to be closed. But several characteristics of the reactor and several silly mistakes by the operators made the experiment dangerous and eventually disastrous.

Chernobyl has boiling-water reactors with graphite moderators. They use the

same neutron-absorbing rods for control and shut down. The control of any reactor is greatly helped by a subtle property of the fission process. A small percentage of the fission neutrons are released after a delay of some seconds, rather than instantaneously. The reactor, in normal operation, has to wait for these delayed neutrons before it can increase power. But if neutron absorbers are withdrawn from the reactor very rapidly and to a sufficient extent, the reactor can go 'prompt critical' and gain power within milliseconds rather than the more usual seconds or minutes. This is still slow by the standards of nuclear explosions (microseconds or less) but it is much faster than the control

cent to 100 times the maximum rated power in about four seconds. The resultant explosive interaction of the over-heated fuel with the cooling water shattered the reactor.

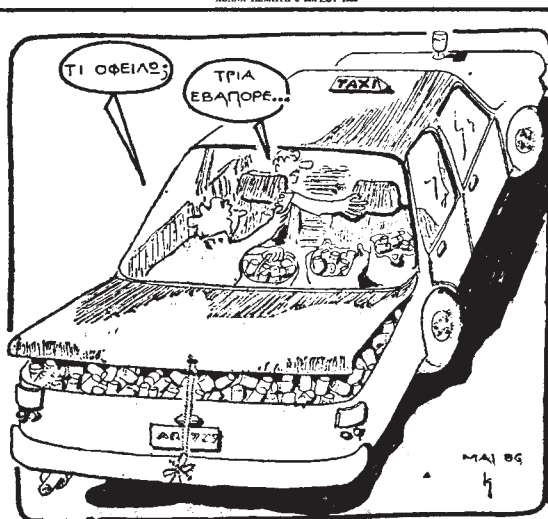
Mould describes this process qualitatively but fails to identify the main causes, especially prompt criticality, and to explain realistically the significance of the operator errors. For example, he notes that the emergency core-cooling system was turned off ("major fault no. 1") but doesn't say that the action was irrelevant to the initiation and development of the accident. He does not explain why the minimum insertion of the control rods was important. He omits the magnitude of the

power surge. Later on, his description of the metabolic differences between iodine, caesium and strontium is confused and misleading.

Perhaps the detailed criticisms can be regarded as technical minutiae, of no concern to the general reader. So how good is the broader picture? The outstanding characteristic of the book is its remarkable coverage of the reported information. It is full of quotations from the Soviet and world press. As an example of data acquisition, it is unchallengeable, and errors, as such, are commendably few. But this is essentially a detective story, so the data have to be analysed, judged for veracity and completeness, and collated. Mould sometimes remarks on the sensationalism of the press reports but, for the most part, he quotes his sources uncritically. His account of the heroism of the fire fighters, for example, is not supplemented by any criticism of the lack of protective clothing

## Η ΚΑΘΗΜΕΡΙΝΗ

ATHINA EPHIMERIS 8 MAI 1986



"How much?" — "Three evaporated milks". An Athens daily newspaper of 8 May 1986 caricatured the local population resorting to panic buying in the wake of Chernobyl.

systems of reactors are designed to handle.

The Chernobyl reactors have two peculiarities. At low power, any increase in steam in the pressure tubes improves the neutron economy and increases the power of the reactor and thus the steam generation — a form of positive feedback. Also, the control rods are shut-off rods. A substantial and specified number of control rods have to be partially in the reactor core so that their movement has an immediate effect. For a variety of reasons, the experiment was started with the reactor at only 7 per cent of its rated power — well into the region of positive feedback — and with very few rods in the core. Automatic control equipment had been disconnected, some of it for no obvious reason. When the steam valve was shut, more steam was generated in the pressure tubes, the power rose, aided by the positive feedback, the control and shut-down system was far too sluggish, and the reactor went prompt critical. A computer simulation, compatible with the physical consequences, suggests that the power went from 7 per

and monitoring.

More gravely, he fails to organize his material coherently. His opening chapter on the power plant fails to identify the weaknesses of design while including extraneous detail. His short chapter on Kiev plunges without warning into the consequences of the previously unmentioned accident. His chapter on evacuation contains the data on the percentage releases from the core and the arrangements for medical follow up; these do not appear in the account of the accident or in the chapter entitled "Follow-up".

The book is liberally illustrated, but the selection of pictures has been uncritical; about half of them could have been omitted without loss. Apart from the chapter on the entombment of the reactor, which is clear and impressive, *Chernobyl* should be treated as an uncorrelated source of (largely) unconfirmed data. This is not "the real story". □

John Dunster, 52 Thames Street, St Ebbes, Oxford OX1 1SU, UK, was formerly Director of the National Radiological Protection Board.



# Pushing beyond the frontiers

David Lindley

**Interactions: A Journey Through the Mind of a Particle Physicist and the Matter of This World.** By Sheldon Glashow with Ben Bova. Warner, New York: 1988. Pp. 297. \$19.95.

ELEMENTARY particle physics fascinates the non-scientist because it tackles simple questions — what is the world made of, and why are there four different forces holding it together? The same appealing simplicity undoubtedly draws scientists of a certain mind to the subject, but their attempts to answer these questions have so far produced a mass of technical detail which tends to obscure the grand design. It is a rare researcher who can spend his days calculating Higgs boson masses or weak coupling angles and still dream at night of the unity of the Universe. Scientists in this venture seem to fall into two sorts; some pioneer the route, solving problems as they arise, while others draw the pieces together and produce the large-scale maps.

To judge from his memoir, Sheldon Glashow is more the pioneer than the mapmaker. His fundamental contributions to understanding the weak interaction and finding how it is connected to electromagnetism are true additions to our knowledge of the world. But having found his territory, Glashow seems disinclined to pause and look around; rather, he moves straight on to the next uncharted spot. As he declares in his final chapter, this puts him in a minority among modern theoretical physicists, many of whom, having staked out their personal vantage point, indulge in lofty and generally fruitless contemplation of the state of the world. But although Glashow's stand for doing science the old-fashioned way may be admirable, it will probably keep him out of the best-seller lists; few casual readers are likely to have the patience to follow him as he recounts the intricate search for neutral currents or bound charm.

The reader's task is made more difficult by the unevenness of the writing, which one is presumably allowed to blame on Glashow's co-author Ben Bova. No effort seems to have been made to pitch the explanations of theories or experimental results at a consistent level. Spontaneous symmetry breaking, for example, is illustrated by recourse to the familiar picture of a marble rolling in the brim of a sombrero, but this is immediately followed with the bald declaration that the shape of the hat is to be thought of as the "potential energy of a quantum field that displays an

internal symmetry". Later, an inadequate account of the Higgs mechanism leads into Glashow's assertion that it is a necessary but ugly part of the standard model of particle physics, like "the flush toilet in a stately mansion". The only purpose for this analogy seems to be to set up a feeble joke about Steven Weinberg.

For those who understand the issues, Glashow's commentary can be useful and illuminating. But those hoping for a more fundamental explanation of what particle physics is about will find only a series of progressively more mystifying and frustrating asides.

The book's general inelegance may be a product of its manufacture. In a collaboration of scientist and professional writer, the ideal function of the second party ought to be to convey the main author's thinking without losing his personal style. The subtitle, *A Journey Through the Mind of a Particle Physicist and the Matter of This World*, reinforces this expectation, but it is hard to avoid the impression that Bova has achieved the opposite. The explanatory sections tend to be confusing and impersonal, while the more personal recollections amount to little more than a series of anecdotes from which one can construct a list of the girlfriends and cars possessed by Glashow in his youth.

But there are times, especially in the later chapters, when Glashow's true enthusiasm for science shines through.

Speaking of his role in the formulation of Grand Unified Theories and of his distaste for the current vogue for superstrings, Glashow emerges as a man with passionate feelings about the state of theoretical physics and its immediate future. That entire physics departments have become beguiled by theories which propose to explain everything, at the cost of never admitting a practical test, is something he clearly finds antithetical to scientific tradition. Too many physicists, he argues, look to the image of Einstein in his later years, trying to understand the Universe by the pure exercise of intellect; most of us should attempt to emulate more reachable heroes.

In most of this book, Glashow does not come across as a contemplative sort. The early chapters may be comparatively plain because he finds it hard to recollect his enthusiasm for problems solved some time ago and now found in the textbooks. Nor does he suffer from the urge to incorporate what he has learned into grand philosophical schemes. He is a solver of puzzles, an inventor of devices and mechanisms, a journeyman scientist. Although such characteristics place him among the majority of working scientists, they may also mean that much of what he has to say is of only incidental interest to those outside science. □

David Lindley is an Assistant Editor on the staff of Nature.

## Surface areas

Robin Williams

**Physics at Surfaces.** By Andrew Zangwill. Cambridge University Press: 1988. Pp. 454. Hbk £40, \$69.50; pbk £15, \$27.95.

WE NOW know a lot about the physics of bulk solids in their crystalline state. Unfortunately, most of the elegant experimental methods that have told us so much about the mechanical, electrical and thermal properties of such solids are quite insensitive to single-atom layers on their surfaces. This is a region where the periodic potential terminates abruptly, and where the crystallographic structure and chemical composition may be very different from those in the bulk. The properties of such layers are often two dimensional in character, again leading to important differences from the bulk behaviour.

Our understanding of the physics of surfaces still lags behind that of solids, but it has come on apace in the past ten years. Good graduate texts on the subject are still in short supply so it is good to see the appearance of Andrew Zangwill's book to complement Prutton's *Surface Physics* (2nd edn 1983) and Woodruff and

Delchar's *Modern Techniques of Surface Science* (1986).

*Physics at Surfaces* is aimed principally at graduate students in physics, physical chemistry and materials science. The first part is concerned with surfaces that are clean on an atomic scale, the second with well-defined layers of atoms and molecules adsorbed onto such surfaces. The author has consciously opted for breadth rather than depth — the book deals with topics ranging from the thermodynamics of surfaces to phase transitions and elementary excitations, and with applied issues ranging from solid-state electronic devices to catalysis.

Inevitably, given the limitation of space, no single topic has been treated rigorously, and the book will therefore be particularly useful as a general text for a broadly based graduate course. It also provides a useful overview for anyone entering the field of surface science; those requiring a more thorough discussion of specific subjects are guided to further reading by a list of references at the end of each chapter. Such is the scope that it is perhaps not surprising that some small errors have crept in, but they do not seriously detract from the value of this useful and timely book.

Robin Williams is a Professor in the Department of Physics, University College Cardiff, PO Box 78, Cardiff CF1 1XC, UK.

# What price space-based interceptors?

David N. Spergel and George B. Field

*The Marshall Institute advocates that the United States should deploy space-based missile interceptors in the near future. At best such weapons would not be cost effective, at worst they would be useless.*

THE deployment of weapons, based on current technologies, as part of the US Strategic Defense Initiative, is a subject of intense debate. A recent report by the Marshall Institute<sup>1</sup>, a non-governmental body which focuses on strategic defence issues, discusses the near-term deployment of a three-layered system. The first layer attacks the intercontinental ballistic missiles (ICBMs) in their boost phase with space-based kinetic kill vehicles (put more simply, space-based interceptors—SBIs). The SBIs would be deployed on satellites in a seven-year programme beginning in 1994. The second (ERIS) layer consists of ground-based kinetic-energy weapons that attack warheads at the end of midcourse flight. The terminal defence uses HEDI missiles to intercept incoming warheads during re-entry.

One of the many questions that must be answered about any such system is: will it be cost effective? Can the Soviets build new missiles that will overwhelm the defensive system at a lower cost than the United States can upgrade its system to meet the new threat? *The President's Strategic Defense Initiative*<sup>2</sup> released by President Reagan on 28 December 1984, states "the defensive system must be able to maintain its effectiveness against the offense at less cost than it would take to develop offensive countermeasures and proliferate the ballistic missiles necessary to overcome it". This 'Nitze criterion' (ref. 3) is incorporated into official policy through National Security Decision Directive 172 (ref. 4).

Here we consider the cost of countering the Soviet deployment of additional missiles, such as the new SS-24s and SS-25s (whose boosters burn for about three minutes) as well as 'fast-burn' boosters (FBBs) with burn times as short as one minute. As SBIs are effective against missiles only during boost phase, deployment of missiles with shorter burn times requires the defence to put more SBIs into orbit. Although the Marshall Institute recognized that missiles of shorter burn time pose a threat to their proposed defensive system<sup>1</sup>, this paper shows that they severely underestimated the number of SBIs required to counter the threat.

Blechman and Utgoff<sup>5</sup> assert that the Soviets will disperse their ICBMs on mobile launchers to ensure that they survive a first strike by the United States. They go on to assume that the launchers will be

dispersed over a large part of the Soviet land mass. Under this assumption, a satellite carrying SBIs is within range of a Soviet booster during a large fraction of its orbit, so relatively few SBIs are needed to counter each new Soviet ICBM in their calculation.

We find that such wide dispersal is not required to ensure survivability. If the Soviets deploy their mobile launchers at a density of one ICBM per 400 km<sup>2</sup>, then ten warheads fired at each launcher would destroy at most 50 per cent of the force<sup>6-8</sup>. The Soviet Union could deploy 1,000 ICBMs in an area of 400,000 km<sup>2</sup> (less than 2 per cent of the area of the Soviet Union) and be certain that a large part of its ICBM force could survive a first strike by the United States.

The Soviet Union will probably deploy its new missiles at a small number of sites, as it does now: *Soviet Military Power*<sup>9</sup> lists three sites for the new SS-25 missiles, four for the SS-19s and six for the SS-18s. We will assume that additional SS-24s, SS-25s and FBBs will be deployed and dispersed over fields of radius 200 km at five different sites to minimize the effectiveness of a SBI boost-phase defence, while allowing the Soviets to attack a range of US targets. In the computer simulations described in Box 1, we find that it is advantageous to the Soviets to locate some of these sites as far north as possible. The Murmansk region at 69° N would be a plausible new site. We shall assume that the Soviets deploy their missiles at bases at latitudes 50° N, 55° N, 60° N, 65° N and 70° N. The calculation is insensitive to the longitude of the bases. As discussed above, 300 ICBMs could be deployed at any of the five sites with little risk to survivability: a US strike with ten warheads attacking each mobile ICBM launcher would expose the average launcher to an overpressure of only 10 psi, because there would be 400 km<sup>2</sup> available to each launcher. Alternatively, the Soviets could defend hardened silos with anti-ballistic missiles (ABMs).

R. M. Gates, the deputy director of the CIA, predicted that "by the mid-1990s, nearly all of the Soviets' currently deployed intercontinental nuclear attack forces—land- and sea-based ballistic missiles and heavy bombers—will be replaced by new and improved systems"<sup>10</sup>. The SBI system is likely to face a modernized Soviet missile force. In 1980 the Soviet Union deployed 250 new ICBMs<sup>11</sup>. For

the purposes of illustration, we assume that the Soviets will respond to the build-up of US missile defences by deploying new ICBMs in the 1990s at comparable rates, and calculate the number of new SBIs needed to counter the 250 new missiles deployed each year.

Several possible Soviet missiles should be considered. In the near term, the Soviets are likely to continue to deploy SS-24 and SS-25 missiles whose burn times are comparable to those of the MX missile (three minutes) rather than older, slower (five-minute) SS-18 missile<sup>9,10</sup>. As the Marshall Institute has noted, the Soviets could well develop and deploy FBBs with even shorter burn times of 120, 90 or even 60 seconds. The Soviet 'Gazelle' ABM demonstrates the feasibility of this technology<sup>11</sup>. Studies by both Martin Marietta<sup>12</sup> and Lockheed<sup>13</sup> suggest that FBBs can be built whose boost-phase ends at an altitude of 80 km with little cost in payload size or missile accuracy. These low-altitude FBBs would evade *all* of the SBIs<sup>11,14</sup>.

The Marshall Institute posits SBIs that accelerate at 20g, achieving a terminal velocity of 6 km s<sup>-1</sup> in 30 seconds. This is the optimal terminal velocity for SBIs (Cunningham *et al.*, unpublished). They assume that these vehicles could achieve a kill probability of 90 per cent. As advocated by the Marshall Institute, we assumed decentralized battle management. Because of random assignment of targets by such a battle management system, achieving the 76 per cent effectiveness envisaged by the Marshall Institute will require a minimum of 1.6 SBIs within range of each missile (see Box 2).

Box 1 describes how we found the optimal configuration for deploying the SBIs. Under the assumption that first-generation SBIs would be unable to track and destroy the post-boost vehicles, which emit no booster flame, we computed the absentee ratio for boost times ranging from 60 to 300 seconds. Table 1 lists these absentee ratios (the ratio of the orbital time to the time that the SBI is in range of its target). We also tabulate the minimum number of SBIs needed to counter the 250 new missiles deployed at the five bases each year. The Marshall Institute estimated that the cost of an SBI, including the cost of launch, its share of the cost of its space platform, and ten years of operations and maintenance, is \$6 mil-



### Box 1 Number of SBIs required

Here we describe our method for calculating the minimum number of SBIs needed to counter new Soviet missiles. The SBI constellation consists of  $N$  orbital families of vehicles. Each orbital family  $k$  ( $k = 1, \dots, N$ ) contains  $M_k$  SBIs distributed in orbits of inclination  $i_k$ .

The ICBMs are deployed in five missile fields of radius  $s$  located at latitude  $\phi_j$  and separated in longitude ( $r_e$  is the radius of the Earth). An SBI in orbit at height  $h_1$ , latitude  $\omega$  and longitude  $\xi$ , relative to missile field  $j$  centred at latitude  $\phi_j$  can destroy an ICBM at a height  $h_2$  above its base only if the distance to the edge of the missile field,

$$D = \{[r_e + h_1]\sqrt{2(1 - \sin \omega \sin \phi_j - \cos \omega \cos \xi \cos \phi_j) - s}\}^2 + (h_2 - h_1)^2\}^{1/2} \quad (1)$$

is less than the range that the SBI can travel during the boost phase,

$$R = v_0 t_{\text{boost}} - \frac{1}{2} \frac{v_0^2}{a} = v_0 \left( t_{\text{boost}} - \frac{1}{2} t_a \right) \quad (2)$$

where  $v_0$  is the SBI's terminal velocity ( $6 \text{ km s}^{-1}$ ),  $a$  is its acceleration ( $20g$ ),  $t_{\text{boost}}$  is the duration of the boost phase, and  $t_a = v_0/a = 30 \text{ s}$  is the time to accelerate the SBI.

The number of SBIs from family  $k$  within range of a missile base  $j$ ,  $L_{jk}$ , is the number of SBIs with  $D(\omega, \xi) < R$ :

$$L_{jk}(\phi_j, M_k, i_k) = \frac{M_k}{2\pi} \int_{-i_k}^{i_k} p(\omega) d\omega \int_{-\pi}^{\pi} d\xi \Theta(D(\omega, \phi_j) - R) \quad (3)$$

where  $p(\omega)$  is the fraction of time that the SBI spends at latitude  $\omega$  and  $\Theta$  is the Heavyside function. The kill-vehicle platform spends equal intervals of time at each longitude, but not at each latitude. Garwin<sup>18</sup> pointed out that the probability density of finding a SBI at a given latitude between  $i$  and  $-i$ ,

$$p(\omega) = \frac{1}{\pi} \frac{\cos \omega}{\sqrt{\sin^2 i - \sin^2 \omega}} \quad (4)$$

is peaked towards  $\omega \approx i$ .

At fixed latitude  $\omega$ , equation (1) implies a maximum relative longitude from which the SVI can reach the missile during its boost phase:

$$\xi_{\text{max}}(\omega, \phi_j) = \cos^{-1} \frac{1 - \frac{[\sqrt{R^2 - (h_2 - h_1)^2 + s}]^2}{2(r_e + h_1)^2} - \sin \omega \sin \phi_j}{\cos \omega \cos \phi_j} \quad (5)$$

Combining equations (1)–(5) yields the number of SBIs in orbital family  $k$  within range of a missile base:

$$L_{jk}(M_k, i_k, \phi_j) = \frac{M_k}{2\pi^2} \int_{-i_k}^{i_k} \int_{-\xi_{\text{max}}}^{\xi_{\text{max}}} \frac{\cos \omega d\omega}{\sqrt{\sin^2 i_k - \sin^2 \omega}} d\xi \\ = \frac{M_k}{\pi^2} \int_{-i_k}^{i_k} \frac{\xi_{\text{max}} \cos \omega(\omega, \phi_j) d\omega}{\sqrt{\sin^2 i_k - \sin^2 \omega}} \quad (6)$$

Horizon effects, which increase the number of SBIs needed to counter each new ICBM (M. Lecar, personal communication) are not included in this calculation.

A Taylor series expansion of equation (6) yields an analytic formula accurate to 1 per cent for  $R \ll r_e$  for SBIs deployed at the optimal inclination attacking missiles from a single base at latitude  $\phi_j$ :

$$L \approx \frac{2.17M}{\pi^2 \sqrt{2} \sin i \cos i} \left( \frac{\sqrt{R^2 - (h_2 - h_1)^2 + s}}{r_e + h_1} \right)^{3/2} \quad (7)$$

Equation (6) was integrated numerically in the calculation of the optimum configuration to defend against an attack from five bases simultaneously.

The total number of missiles within range of each base is just  $\sum_{k=1}^N L_{jk}(M_k, i_k, \phi_j)$ . Box 2 shows that there must be at least 79 SBIs within range of the 50 missiles at each base if leakage is to be limited to 24 per cent.

The problem of finding which distribution of  $M_k$  SBIs over  $N$  inclinations  $i_k$  minimizes the total number of SBIs ( $\sum_{k=1}^N M_k$ ) subject to the constraint that there must be at least 79 SBIs within range of each missile ( $\sum_{k=1}^N L_{jk}(M_k, i_k, \phi_j) > 79$ ) is a multi-dimensional minimization subject to constraints. We adapted the AMOEBA algorithm<sup>19</sup> to the problem and numerically determined the solution. The calculations assumed a satellite altitude of 400 km, a base radius of 200 km and burn-out heights were taken from Fig. 4 of ref. 13. Table 2 shows the optimum SBI configuration for attacking missiles from a given base destroyed by each orbital family of SBIs.

### Box 2 Decentralized battle management

The Marshall Institute report<sup>1</sup> advocates a decentralized battle management system, which allows a decision to be made independently by each SBI platform as to which missiles it will attack. This system has the advantage of reducing the complexity of the battle management system, but has the disadvantage of reducing system effectiveness.

The best that can be achieved with decentralized management is a random assignment of SBIs in each cluster to each missile. This implies that if there are  $I$  SBIs within range of  $J$  missiles, the probability of having  $m$  SBIs assigned to a given missile is

$$p(m) = \exp(-x) \frac{x^m}{m!} \quad (8)$$

where  $x = I/J$ . Suppose each SBI has a probability  $k$  of

destroying a missile. If the SBIs fail independently, then the probability that a given missile survives  $m$  SBIs is  $(1-k)^m$ . We multiply this factor by the probability of having  $m$  SBIs attack a given missile and sum over  $m$  to find the fraction of missiles that leaks through the boost phase defence layer:

$$f_{\text{leakage}} = \sum_{m=0}^{\infty} \exp(-x) \frac{x^m}{m!} (1-k)^m = \exp(-kx). \quad (9)$$

A leakage rate  $f_{\text{leakage}}$  necessitates a minimum of  $x$  SBIs per satellite, where  $x = \log(f_{\text{leakage}})/k$ .

The Marshall Institute report assumes that 76 per cent of the Soviet missiles are destroyed by SBIs whose kill probability  $k$  is 0.9. The corresponding 0.24 leakage rate requires a minimum of 1.6 SBIs within range of each Soviet missile. If the failures of SBIs are correlated, more SBIs will be needed.

lion. Using this estimate, we computed the annual US expenditure needed to counter the 250 missiles. The penultimate paragraph explains why the figures listed in the last column of Table 1 are conservative cost estimates.

Estimates for Soviet missile costs vary. If the price of an SS-25 is comparable to that of a mass-produced MX missile (\$50 million including operations and maintenance<sup>15</sup>), then 250 new SS-25s will cost the Soviet Union \$12.5 billion. Table 1 indicates that the United States must spend \$42 billion to counter the new threat, resulting in a cost exchange ratio of 3:1 in favour of the Soviet Union. If the Soviets move to FBBs, the cost exchange ratio becomes even more advantageous for the offence. Estimates for the cost of an FBB (including launcher) vary from \$10 million<sup>13</sup> up to \$90 million<sup>11</sup>, so that the Soviet Union could add 250 new missiles to its missile force each year for \$2.5–22.5 billion. If the Soviets deployed missiles with a 90-second burn time, the United States would have to spend \$180 billion annually for the SBIs required to counter this new force, resulting in a cost exchange ratio of 8:1 to 70:1 in favour of the Soviet Union. Soviet missiles with 60-second burn time would be invulnerable to attack during the boost phase<sup>11</sup>. They will however, have to coast for about 15 seconds before warheads and decoys can be deployed from the post-boost vehicle (PBV)<sup>16</sup>. Even if the PBV can be attacked in this coast phase, the cost exchange ratio is 13:1 to 120:1 (depending on the costs of the FBBs) in favour of the Soviet Union.

In making these estimates, we made several assumptions that were favourable to the defence. We accepted the Marshall Institute estimate that SBIs would be 90 per cent effective in killing their assigned targets. If a more realistic estimate of 50 per cent effectiveness is adopted, then the number of SBIs needed to counter the Soviet threat would increase by 80 per cent. We ignored the possibility of counter-measures that would enable the booster to evade many of the SBIs. If the Soviets located all of their missiles at one site, the absentee ratio would increase by a factor of three, trebling the cost of the defensive system. We assumed that the interceptors would be fired immediately after the launch of the missile. Any delay due to clouds or system start-up time would increase the absentee ratio and therefore raise the system cost. The American Physical Society study<sup>16</sup> estimates a 20–30-second system start-up time. We assumed that the SBIs could only attack booster rockets, because current technology can track the booster flame emitting  $8 \text{ MW sr}^{-1}$ . Detecting PBVs emitting  $2 \times 10^{-6}$  the luminosity of a booster ( $\sim 30 \text{W sr}^{-1}$ ) against an Earth background may prove impossible<sup>16</sup>. If sufficiently sensitive sensors could be developed that

Table 1 Minimum cost of countering Soviet buildup

| Soviet missile 'burn time' | Absentee ratio | No. of SBIs needed to counter 250 missiles | Cost of new SBIs needed to counter 250 missiles (\$ billion) |
|----------------------------|----------------|--------------------------------------------|--------------------------------------------------------------|
| 60 s                       | —              | —                                          | —                                                            |
| 60 s + 15 s coast          | 121            | 48,000                                     | 300                                                          |
| 90 s                       | 76             | 30,000                                     | 180                                                          |
| 120 s                      | 38             | 15,000                                     | 90                                                           |
| 150 s                      | 24             | 9,600                                      | 58                                                           |
| 180 s                      | 18             | 7,000                                      | 42                                                           |
| 300 s                      | 5              | 2,100                                      | 13                                                           |

Table 2 Optimum configuration of SBIs (180 s burn time)

| Orbital inclination ( $i_k$ ) | No. of SBIs ( $M_k$ ) | Number of missiles destroyed ( $L_{jk}$ ) |       |       |       |       |
|-------------------------------|-----------------------|-------------------------------------------|-------|-------|-------|-------|
|                               |                       | Base latitude ( $\phi_j$ )                |       |       |       |       |
|                               |                       | 50° N                                     | 55° N | 60° N | 65° N | 70° N |
| 76.9° N                       | 1,533                 | 32                                        | 22    | 17    | 14    | 12    |
| 74.0° N                       | 1,902                 | 37                                        | 34    | 23    | 18    | 16    |
| 69.5° N                       | 681                   | 10                                        | 12    | 11    | 7     | 6     |
| 57.6° N                       | 1,779                 | 0                                         | 8     | 20    | 26    | 28    |
| 56.7° N                       | 1,118                 | 0                                         | 3     | 11    | 16    | 17    |
|                               |                       | 79                                        | 79    | 82    | 81    | 79    |

would enable SBIs to attack post-boost vehicles, then the absentee ratio would decrease and this would lower system cost. We assumed that the SBI system could defend itself against attack and ignored the start-up costs associated with the system. We also ignored the cost of improving the other layers to handle the additional threat posed by the new warheads that leak through the boost-phase layer.

We conclude that the boost-phase defence advocated by the Marshall Institute would not be cost effective against SS-24s or SS-25s. It would be even less cost effective against fast-burn boosters and totally ineffective against boosters with burn times of less than 65 seconds. Its deployment could result in an arms race that is economically unfavourable to the United States. Simple geometrical constraints pose a serious problem for any boost-phase defence composed of SBIs. Two recent studies<sup>10,17</sup> agree with the conclusion of this paper that the absentee ratio for SBIs attacking missiles with short boost times is very large. Furthermore, matching Soviet deployment of SS-24s, SS-25s and FBBs will be very costly.

We thank M. Lecar for many useful discussions and M. Brower, A. Carter, S. Drell, P. Duffy, R. Garwin, J. Goodman, G. Lewis and E. Salpeter for helpful comments on an earlier manuscript.

1. Report of the Technical Panel on Deployment of Missile Defense in the 1990s (George C. Marshall Institute, Washington, DC, 1987).

2. The President's Strategic Defense Initiative (Government Printing Office, Washington, DC, January 1985).
3. Nitze, P. H. in *Strategic Defense Initiative: Folly or Future* (eds Haley, P. E. & Merritt, J.) 37–43 (Westview Press, Boulder, 1987).
4. The Strategic Defense Initiative (Spec. Rep. No. 129) (US Department of State Bureau of Public Affairs, Washington, DC, 1985).
5. Blechman, B. M. & Utgoff, V. A. *Fiscal and Economic Implications of Strategic Defences* (School of Advanced International Studies Papers in International Affairs No. 12) (Westview Press, Boulder, 1986).
6. Cochran, T. B., Arkin, W. M. & Hoenig, M. M. *Nuclear Weapon Databook, Volume I: Nuclear Forces and Capabilities* (Ballinger, Cambridge, Massachusetts, 1984).
7. Gallatou, S. & Dolan, P. J. *The Effects of Nuclear Weapons*, 3rd edn (US Department of Defense, Washington, DC, 1977).
8. Smith, B. A. *Aviat. Week Space Technol.* p. 54, 6 Feb. 1984.
9. *Soviet Military Power*, 5th edn (Government Printing Office, Washington, DC, 1986).
10. *Ballistic Missile Defense Technology* (Government Printing Office, Washington, DC, September 1985).
11. Brower, M. *Nucleus*, 9(2), 1 (1987).
12. Short Burn-time ICBM Characteristics and Considerations (Martin Marietta Denver Aerospace, Denver, Colorado, 20 July 1983).
13. Deitz, B. *Fast-Burn Missile Requirements Analysis* (Lockheed Missiles and Space, Sunnyvale, California, 1986).
14. Carter, A. B. *Directed Energy Missile Defense in Space* (Background pap. for Office of Technology Assessment, Washington, DC, April 1984).
15. Field, G. B. & Spergel, D. N. *Science* **231**, 1387 (1986).
16. Report to the APS of the Study Group on Science and Technology of Directed Energy Weapons, *Rev. Mod. Phys.* **59**, S1 (1987).
17. Cunningham, C., *Energy and Technical Review* 16 (Lawrence Livermore Laboratory, July 1987).
18. Garwin, R. *Nature* **315**, 286 (1985).
19. Press, W. H., Flannery, B. P., Teukolsky, A. & Vetterling, W. T. *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, UK, 1986).

David N. Spergel is in the School of Natural Science, Institute for Advanced Study, Princeton, New Jersey 08540, USA. George B. Field is in the Department of Astronomy, Harvard University, Cambridge, Massachusetts 02138, USA.



# Human basophil degranulation triggered by very dilute antiserum against IgE

E. Davenas, F. Beauvais, J. Amara\*, M. Oberbaum\*, B. Robinzon†, A. Miadonna‡, A. Tedeschi‡, B. Pomeranz§, P. Fortner§, P. Belon, J. Sainte-Laudy, B. Poitevin & J. Benveniste||

INSERM U 200, Université Paris-Sud, 32 rue des Carnets, 92140 Clamart, France

\* Ruth Ben Ari Institute of Clinical Immunology, Kaplan Hospital, Rehovot 76100, Israel

† Department of Animal Sciences, Faculty of Agriculture, PO Box 12, The Hebrew University of Jerusalem, Rehovot 76100, Israel

‡ Department of Internal Medicine, Infectious Diseases and Immunopathology, University of Milano, Ospedale Maggiore Policlinico, Milano, Italy

§ Departments of Zoology and Physiology, Ramsay Wright Zoological Laboratories, University of Toronto, 25 Harbord Street, Toronto, Ontario M5S 1A1, Canada

|| To whom correspondence should be addressed.

*When human polymorphonuclear basophils, a type of white blood cell with antibodies of the immunoglobulin E (IgE) type on its surface, are exposed to anti-IgE antibodies, they release histamine from their intracellular granules and change their staining properties. The latter can be demonstrated at dilutions of anti-IgE that range from  $1 \times 10^3$  to  $1 \times 10^{120}$ ; over that range, there are successive peaks of degranulation from 40 to 60% of the basophils, despite the calculated absence of any anti-IgE molecules at the highest dilutions. Since dilutions need to be accompanied by vigorous shaking for the effects to be observed, transmission of the biological information could be related to the molecular organization of water.*

THE antibodies responsible for human immediate hypersensitivity belong to the IgE isotype<sup>1</sup>. The most salient feature of IgE is its capacity to bind to mast cell and polymorphonuclear basophil membranes through receptors with high affinity<sup>2</sup>. Human basophils are specifically challenged by immunological stimuli such as allergens or anti-IgE antiserum that can bridge IgE molecules in membrane<sup>3</sup>. This process triggers transmembrane and intracellular signals followed by granule exocytosis with the release of histamine and loss of metachromatic staining of basophil granules by a basic dye such as toluidine blue. Optical basophil degranulation is well correlated with other *in vitro* and *in vivo* procedures for the diagnosis of allergy<sup>4,7</sup>.

In preliminary experiments, degranulation of human basophils contained in leukocyte suspensions was induced not only by the usual concentration of anti-IgE antibody ( $1 \times 10^3$  dilution of anti-IgE antiserum, corresponding to  $2.2 \times 10^{-16}$  M anti-IgE antibody in the assay), but also by very low concentrations of this antibody ( $2.2 \times 10^{-16/18}$  M), where the number of IgG anti-IgE molecules in the assay is supposedly too low to trigger the process. We then further explored this phenomenon.

Serial tenfold dilutions of goat anti-human IgE (Fc) antiserum (1 mg specific antibody per ml) were prepared in HEPES-buffered Tyrode's solution containing human serum albumin (HSA) down to  $1 \times 10^{60}$  dilution, corresponding to a  $2.2 \times 10^{-16}$  M theoretical concentration (th) in the assay (see Fig. 1 legend for methods). The expected basophil degranulation, which was assessed by counting cells with metachromatical properties, was observed after exposure of leukocyte preparations to low antiserum dilutions with a maximum at  $\sim 1 \times 10^3$

dilution. Successive peaks of degranulation varying between 40 and 60% were then found down to  $1 \times 10^{60}$  dilution, with periods of 6 to 9 tenfold dilutions (Fig. 1a). In other experiments, the antiserum was serially diluted a hundred-fold down to  $1 \times 10^{120}$  (to give  $2.2 \times 10^{-126}$  M th in the assay) and similar results were obtained (Fig. 1b). Degranulation induced by high dilutions of anti-IgE antiserum was observed in ten experiments on the full range of dilutions down to  $1 \times 10^{60}$ , when at least 70 similar results were obtained at one or the other part of the high dilution scale in the participating laboratories (Toronto, preliminary results). As controls, goat antihuman IgG (Fc) antiserum (Fig. 1b,  $n = 4$ ) or Tyrode's solution containing HSA ( $n = 5$ ) were diluted down to  $1 \times 10^{120}$  and  $1 \times 10^{30}$ , respectively. Cells incubated in conditions identical to those with anti-IgE antiserum gave no significant degranulation. The repetitive waves of anti-IgE-induced degranulation were reproducible, but the peaks of degranulation could shift by one or two dilutions with every fresh sequential dilution of anti-IgE and depended on the blood sample. The waves of basophil degranulation were also seen with substances other than anti-IgE anti-serum at high and low dilutions, such as monoclonal anti-human IgE antibodies, specific antigen in allergic patients or in peroxidase-immunized rabbits, phospholipase A<sub>2</sub> from bee venom or porcine pancreas, the Na<sup>+</sup> ionophore monensin (up to 90% degranulation at  $1 \times 10^{-30}$  M th) and the Ca<sup>2+</sup> ionophores A23187 and ionomycin ( $1 \times 10^{-38}$  M th). The specificity of the observed effects at high dilutions (already noted when comparing antiserum against IgE with antiserum against IgG) was further strikingly illustrated in the ionophore experiments, because removing the corresponding ion from the cellular environment blunted basophil

**Table 1** Basophil counts after exposure to anti-IgE antiserum at low and high dilutions

| Samples                 | Experiment 1     | Experiment 2    | Experiment 3    | Experiment 4    |
|-------------------------|------------------|-----------------|-----------------|-----------------|
| Tyrode's-HSA*           | 81.3 ± 1.2†      | 89.0 ± 3.1      | 81.7 ± 2.2      | 106.7 ± 1.8     |
| Tyrode's-HSA            | 81.6 ± 1.4       | 87.7 ± 1.4      | 83.0 ± 1.0      | 105.0 ± 1.2     |
| Tyrode's-HSA            | 80.0 ± 1.5       | 88.0 ± 2.3      | 81.7 ± 1.8      | 105.7 ± 0.9     |
| aIgE $1 \times 10^{18}$ | 35.5 ± 1.8 (56)‡ | 42.3 ± 4.8 (53) | 27.7 ± 0.7 (66) | 40.0 ± 1.5 (62) |
| aIgE $2 \times 10^{32}$ | 77.6 ± 0.8 (4)   | 87.3 ± 1.2 (3)  | 66.3 ± 2.3 (18) | 93.7 ± 1.9 (12) |
| aIgE $1 \times 10^{33}$ | 76.0 ± 1.1 (6)   | 88.7 ± 1.8 (1)  | 77.7 ± 1.8 (4)  | 74.7 ± 2.8 (30) |
| aIgE $1 \times 10^{34}$ | 53.6 ± 1.4 (33)  | 52.7 ± 1.4 (41) | 38.0 ± 0.6 (53) | 48.3 ± 2.4 (55) |
| aIgE $1 \times 10^{35}$ | 45.0 ± 0.5 (44)  | 35.0 ± 1.0 (61) | 41.3 ± 1.8 (49) | 49.3 ± 1.2 (54) |
| aIgE $1 \times 10^{36}$ | 49.0 ± 1.7 (40)  | 50.3 ± 0.7 (44) | 55.0 ± 2.1 (32) | 74.3 ± 2.3 (31) |
| aIgE $1 \times 10^{37}$ | 79.0 ± 2.3 (2)   | 85.3 ± 0.7 (5)  | 73.3 ± 1.7 (10) | 105.3 ± 0.7 (0) |

Blind experiments: test tubes were randomly coded twice by two independent pairs of observers and assayed. The codes were simultaneously broken at the end of all experiments. Dilutions of anti-IgE antiserum were performed as described in legend to Fig. 1.

\* Uncoded additional tubes for negative (Tyrode's-HSA) or positive (aIgE  $1 \times 10^{-3}$ ) controls. † Data represent the mean ± s.e. of basophil number actually counted in triplicate (see legend to Fig. 1 for methods). ‡ Number in parenthesis indicates percentage degranulation compared with Tyrode's-HSA.

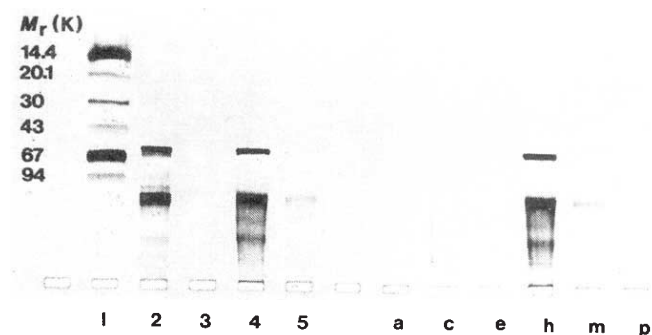
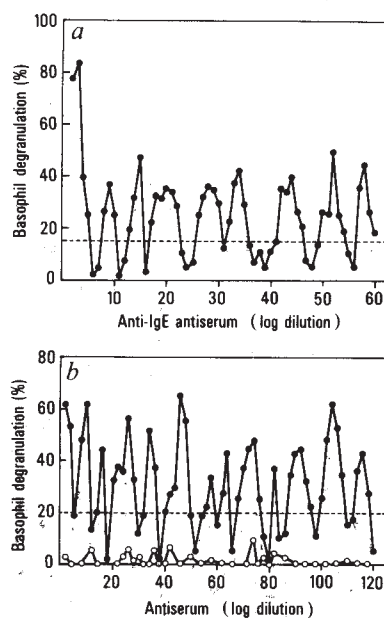
**Fig. 1** Human basophil degranulation induced either by anti-IgE anti-serum (●) diluted tenfold from  $1 \times 10^2$  down to  $1 \times 10^{60}$  (a) or hundredfold down to  $1 \times 10^{120}$  (b) or by anti-IgG anti-serum (○) diluted hundredfold from  $1 \times 10^2$  down to  $1 \times 10^{120}$  (representatives of at least 10 experiments for anti-IgE and 4 experiments for anti-IgG). The significant ( $P < 0.05$ ) percentage of degranulation was 15% (a) and 20% (b). (....) relation to the number of counted basophils from control wells<sup>15</sup>.

**Methods** Goat anti-human IgE (Fc) antiserum or as a control, goat anti-human IgG (Fc) antiserum (Nordic Immunology, The Netherlands) was serially diluted as indicated above in HEPES-buffered Tyrode's solution (in g l<sup>-1</sup>: NaCl, 8; KCl, 0.195; HEPES, 2.6; EDTA-Na<sub>2</sub>, 1.040; glucose, 1 human serum albumin (HSA), 1.0; heparin, 5000 U per l; pH 7.4). Between each dilution, the solution was thoroughly mixed for 10 s using a Vortex. Given the molecular weight of IgG molecules (150,000), the  $1 \times 10^{60}$  and  $1 \times 10^{120}$  dilutions correspond in the assay to  $2.2 \times 10^{-66}$  M (th) and  $2.2 \times 10^{-120}$  M (th) respectively. Venous blood (20 ml) from healthy donors was collected using heparin (1 U per ml) and a mixture of 2.5 mM EDTA-Na<sub>2</sub>/2.5 mM EDTA-Na<sub>2</sub> (final concentrations) as anticoagulants and allowed to sediment. The leukocyte-rich plasma was recovered, twice washed by centrifugation (400g, 10 min) and finally resuspended in an aliquot of HEPES-buffered Tyrode's solution. The cell suspension (10  $\mu$ l) was deposited on the bottom of each well of a microtitre plate containing 10  $\mu$ l CaCl<sub>2</sub> (5 mM final) and 10  $\mu$ l of either of anti-IgE or anti-IgG antiserum dilutions. To a control well were added 10  $\mu$ l CaCl<sub>2</sub> and 10  $\mu$ l Tyrode's but no anti-IgE or anti-IgG antiserum. Plates were then incubated at 37°C for 30 min. Staining solution (90 ml: 100 mg toluidine blue and 280  $\mu$ l glacial acetic acid in 100 ml 25% ethanol, pH 3.2 — 3.4) was added to each well and the suspension thoroughly mixed. Specifically restained basophils (non-degranulated basophils) were counted under a microscope using a Fuchs-Rosenthal haemocytometer. The percentage of basophil degranulation was calculated using the following formula: Basophil no. in control — basophil no. in sample/ basophil no. in control  $\times 100$ . Between 60 and 120 basophils were counted in cell suspensions from control wells after incubation either in the absence of anti-IgE antiserum, or in the presence of anti-IgG antiserum.

degranulation.

To confirm these surprising findings, four blind experiments were carried out (Table 1). In all cases the results were clear-cut, with typical bell-shaped degranulations at anti-IgE dilutions from  $1 \times 10^{32}$  to  $1 \times 10^{37}$ . The replicates were usually very close and of high significance (ANOVA test). In a fifth experiment, 7 control tubes and 3 tubes containing a dilution previously determined as active ( $1 \times 10^{34}$ ) were counted blind: basophil degranulation was  $7.7 \pm 1.4\%$  for the controls, and 44.8, 42.8 and 45.7% for the tubes containing diluted anti-IgE. The random chance in all these experiments was 2% and therefore the cumulative results statistically confirm the measured effect.

Two further blind experiments were performed using the usual dilution procedure: of the 12 tubes used in the first experiment (Table 2), 2 tubes contained goat anti-human antiserum IgE at  $1 \times 10^2$  and  $1 \times 10^3$  dilutions, 6 tubes contained dilutions from  $1 \times 10^{32}$  to  $1 \times 10^{37}$ , and 4 tubes buffer-HSA alone. The tubes were then randomly coded twice by three parties, one of which kept the two codes. The 12 tubes were each divided into 4. Three batches of 12 tubes were lyophilized, one of which was used for gel electrophoresis, one for assay of monoclonal antibodies, and the last (with the unlyophilized sample) for gel electrophoresis and basophil degranulation. By comparing the results of the different tests it was easy to identify the tubes containing IgE at normal concentrations compared with the tubes containing highly diluted IgE and the control tubes. When the codes were broken, the actual results exactly fitted those predicted, but HSA and its aggregates were present in all solutions and complicated interpretation of the gel electrophoresis.



**Fig. 2** Electrophoresis (polyacrylamide 7–15%, bands revealed by silver staining): samples numbered 1 to 5 are standards for the blind experiments a, c, e, h, m, p. Lane 1, Molecular weight standards for electrophoresis; lane 2, monoclonal IgG added with human serum albumin; lane 3, Tyrode's buffer without human serum albumin; lane 4,  $1 \times 10^2$  anti-IgE dilution; lane 5,  $1 \times 10^3$  dilution. Samples tested blind: a and c, buffer; e,  $1 \times 10^{36}$  anti-IgE dilution; h,  $1 \times 10^{32}$  anti-IgE dilution; m,  $1 \times 10^{37}$  anti-IgE dilution; p,  $1 \times 10^{35}$  anti-IgE dilution.

So we performed another almost identical experiment, using 6 tubes containing unlyophilized samples and buffer without HSA. Four tubes contained antibody at  $1 \times 10^2$ ,  $1 \times 10^3$ ,  $1 \times 10^{35}$  and  $1 \times 10^{36}$  dilutions, and 2 contained buffer alone. These tubes were coded and assayed according to the above protocol. The decoded results were clear-cut, high basophil degranulation being obtained with  $1 \times 10^2$ ,  $10^3$ ,  $10^{35}$  and  $10^{36}$  dilutions, but no anti-IgE activity or immunoglobulins were detected either in the control tubes or in assays containing the  $1 \times 10^{35}$  and  $10^{36}$  dilutions (Tables 2 and 3 and Fig. 2). Thus there is no doubt that there was basophil degranulation in the absence of any detectable anti-IgE molecule.

These results may be related to the recent double-blind clinical study of Reilly *et al.*<sup>8</sup> which showed a significant reduction of symptoms in hay-fever patients treated with a high dilution ( $1 \times 10^{60}$ ) of grass pollen versus placebo, and to our *ex vivo* experiments in the mouse<sup>9</sup>. We have extended these experiments to other biological systems: using the fluorescent probe fura-2, we recently demonstrated changes in intracellular Ca<sup>2+</sup> levels in human platelets in the presence of the Ca<sup>2+</sup> ionophore ionomycin diluted down to  $1 \times 10^{-39}$  M th (F. B. *et al.*, unpublished results).

Using the molecular weight of immunoglobulins and Avogadro's number, we calculate that less than one molecule of antibody is present in the assay when anti-IgE antiserum is diluted to  $1 \times 10^{14}$  (corresponding to  $2.2 \times 10^{-20}$  M). But in the experiments reported here we have detected significant basophil degranulation down to the  $1 \times 10^{120}$  dilution. Specific effects have also been triggered by highly diluted agents in other *in vitro* and *in vivo* biological systems<sup>8–11</sup>, but still remain unexplained. The valid use of Avogadro's number could be questioned, but we are dealing with dilutions far below the Avogadro limit ( $1 \times 10^{60}$  and below). It could be argued that our serial dilution procedure is subject to experimental error, but this is ruled out because: (1) pipette tips and glass micro-pipettes were discarded between each dilution (performed under laminar flow hood). (2) The c.p.m. in tubes containing serially diluted radioactive compounds decreased in proportion to the degree of dilution down to the background (data not shown). (3) Contamination would not explain the successive peaks of activity that evoke a periodic phenomenon and not a monotonous dose–effect curve, as usually observed when concentration of an agonist decreases. (4) To eliminate the possibility of contaminating molecules present in the highly diluted solutions, we carried out two series of experiments which can be summarized as follows. An Amicon membrane with molecular weight cut-off 10K retained the basophil degranulating IgG (150K) present at low dilutions ( $1 \times 10^2$ ,  $1 \times 10^3$ ) in anti-IgE antiserum. By contrast, the activity present at high dilutions ( $1 \times 10^{37}$ ,  $1 \times 10^{39}$ ) was totally recovered in the 10K Amicon filtrate. Anion or cation exchange chromatography,



**Table 2** Comparison of basophil degranulation with the presence of immunoglobulins and anti-IgE activity in dilutions performed in HSA-containing Tyrode's

| Samples                   | Basophil degranulation (%) <sup>*</sup> |    |     | Gel electrophoresis <sup>†</sup> |    | Anti-IgE activity<br>$\mu\text{ml}^{-1}$ |
|---------------------------|-----------------------------------------|----|-----|----------------------------------|----|------------------------------------------|
|                           | I                                       | II | III | A                                | B  |                                          |
| Tyrode's-HSA              | 0                                       | 0  | 0   | —                                | —  | $< 1 \times 10^{-3}$                     |
| Tyrode's-HSA              | 0                                       | 0  | 0   | —                                | —  | $< 1 \times 10^{-3}$                     |
| Tyrode's-HSA              | 0                                       | 0  | 0   | —                                | —  | $< 1 \times 10^{-3}$                     |
| Tyrode's-HSA              | 0                                       | 0  | 0   | —                                | —  | $< 1 \times 10^{-3}$                     |
| alGE $1 \times 10^{-2}$ ‡ | 53                                      | 50 | 33  | ++§                              | ++ | ND                                       |
| alGE $1 \times 10^{-2}$   | 51                                      | 44 | 37  | ++                               | ++ | 10.6                                     |
| alGE $1 \times 10^{-3}$   | 65                                      | 38 | 45  | ++                               | —  | 1.1                                      |
| alGE $1 \times 10^{-32}$  | 7                                       | 26 | 22  | —                                | —  | $< 1 \times 10^{-3}$                     |
| alGE $1 \times 10^{-33}$  | 37                                      | 0  | 13  | —                                | —  | $< 1 \times 10^{-3}$                     |
| alGE $1 \times 10^{-34}$  | 45                                      | 37 | 20  | —                                | —  | $< 1 \times 10^{-3}$                     |
| alGE $1 \times 10^{-35}$  | 39                                      | 41 | 34  | —                                | —  | $< 1 \times 10^{-3}$                     |
| alGE $1 \times 10^{-36}$  | 31                                      | 29 | 39  | —                                | —  | $< 1 \times 10^{-3}$                     |
| alGE $1 \times 10^{-37}$  | 23                                      | 12 | 29  | —                                | —  | $< 1 \times 10^{-3}$                     |

Blind experiments and dilution protocols as in Table 1. —, Lack of strained bands. ND, not determined. A faint band corresponding to IgG appeared after reduction by 2-mercaptoethanol.

\* Basophil degranulation tests I, II, III were performed using 3 different blood samples (see Fig. 1). Percentage basophil degranulation induced by alGE, as compared to Tyrode's HSA, was calculated from duplicates.

† Electrophoresis (polyacrylamide 7–15%, revealed by silver staining) was carried out in Rehovot (A) and at INSERM U 200 (B).

‡ Uncoded additional tube for positive control.

§ ++, + Bands correspond to IgG present in large or small amounts.

according to the type of resin used and the pH, did or did not retain the anti-IgE IgG at low dilutions, whereas the same activity at high dilution was always excluded from the columns and fully recovered in the first eluate. These filtration and ion-exchange experiments demonstrated that the activity of the antiserum at high dilution cannot result from contamination of the highly diluted solution with the starting material. They showed, in addition, that the high-dilution activity does not present in space the steric conformation of an IgG molecule as it acts like a 150K charged molecule, but is not retained by the 10K filter or by a charged chromatography column.

We then investigated the physical chemical nature of the entity active at high dilution. Our results can be summarized as follows. (1) The importance of agitation in the transmission of information was explored by pipetting dilutions up and down ten times and comparing with the usual 10-s vortexing. Although the two processes resulted in the same dilution (degranulations at  $1 \times 10^2$  and  $1 \times 10^3$  were superimposable whatever the dilution process), degranulation did not occur at high dilution after pipetting. Ten-second vortexing was the minimum time required, but vortexing for longer (30 or 60 s) did not increase high-dilution activity. So transmission of the information depended on vigorous agitation, possibly inducing a sub-molecular organization of water or closely related liquids. (2) The latter is possible as ethanol and propanol could also support the phenomenon. In contrast, dilutions in dimethylsulphoxide did not transmit the information from one dilution to the other, but increasing the proportion of water in dimethylsulphoxide resulted in the appearance and increment of the activity at high dilutions. (3) Heating, freeze-thawing or ultrasonication suppressed the activity of highly diluted solutions, but not the activity of several active compounds at high concentrations. A striking feature was that molecules reacted to heat according to their distinctive heat sensitivity, whereas all highly diluted solutions ceased to be active between 70 and 80°C. This result suggests a common mechanism operating at high dilution, independent of the nature of the starting molecule.

Therefore we propose that none of the starting molecules is present in the dilutions beyond the Avogadro limit and that specific information must have been transmitted during the dilution/shaking process. Water could act as a 'template' for the molecule, for example by an infinite hydrogen-bonded network<sup>12</sup>, or electric and magnetic fields<sup>13,14</sup>. At present we can only speculate on the nature of the specific activity present in the highly diluted solutions. We can affirm that (1) this activity was established under stringent experimental conditions, such as

**Table 3** Comparison of basophil degranulation with the presence of immunoglobulins and anti-IgE activity in dilutions performed in Tyrode's without HSA.

| Samples                   | Basophil degranulation (%) |    | Gel electrophoresis |    | Anti-IgE activity<br>$(\mu\text{ml}^{-1})$ |
|---------------------------|----------------------------|----|---------------------|----|--------------------------------------------|
|                           | I                          | II | A                   | B  |                                            |
| Tyrode's                  | 0                          | 0  | —                   | —  | $< 1 \times 10^{-3}$                       |
| Tyrode's                  | 0                          | 0  | —                   | —  | $< 1 \times 10^{-3}$                       |
| alGE $1 \times 10^{-2}$ * | 85                         | 48 | ++                  | ++ | ND                                         |
| alGE $1 \times 10^{-2}$   | 81                         | 47 | ++                  | ++ | 32.6                                       |
| alGE $1 \times 10^{-3}$ * | ND                         | ND | +                   | +  | ND                                         |
| alGE $1 \times 10^{-3}$   | 75                         | 53 | +                   | +  | ND                                         |
| alGE $1 \times 10^{-35}$  | 35                         | 31 | —                   | —  | $< 1 \times 10^{-3}$                       |
| alGE $1 \times 10^{-36}$  | 40                         | 35 | —                   | —  | $< 1 \times 10^{-3}$                       |

\* Uncoded tubes for positive control of basophil degranulation and/or gel electrophoresis.

ND, not determined.

blind double-coded procedures involving six laboratories from four countries; (2) it is specific for the ligand first introduced, as illustrated when goat antiserum (IgG) anti-human IgE, but not goat IgG anti-human IgG supported this phenomenon. The link between high and low anti-IgE dilutions is shown as we could not detect basophil degranulation at high dilutions if it did not occur within the classical range. High dilutions of histamine, but not of its carboxylated precursor histidine, inhibited IgE-dependent basophil degranulation. Finally, ionophores at high dilution did not work when the specific ion was removed from the cell suspension (F.B., unpublished results). (3) Using six biochemical and physical probes, we demonstrated that what supports the activity at high dilutions is not a molecule. (4) Whatever its nature, it is capable of 'reproducing' subtle molecular variations, such as the rearrangement of the variable region of an IgG (anti- $\epsilon$  versus anti- $\gamma$ ) molecule.

The precise nature of this phenomenon remains unexplained. It was critical that we should first establish the reality of biological effects in the physical absence of molecules. The entities supporting this 'metamolecular' biology can only be explored by physical investigation of agitation causing interaction between the original molecules and water, thus yielding activity capable of specifically imitating the native molecules, though any such hypothesis is unsubstantiated at present.

We thank Professor Z. Bentwich from Ruth Ben Ari Institute for supervision of experiments conducted in Rehovot. The participation of J. Geen (Univ. Toronto), B. Descours and C. Hieblot (INSERM U 200) in experiments and of V. Besso in editing is gratefully acknowledged. This work is dedicated to the late Michel Aubin, who played a decisive role in initiating it.

Received 24 August 1987; accepted 13 June 1988.

1. Ishizaka, K., Ishizaka, T. & Hornbrook M. M. *J. Immun.* **97**, 75–85 (1966).
2. Metzger, H. *et al. A. Rev. Immun.* **4**, 419–470 (1986).
3. Ishizaka, T., Ishizaka, K., Conrad, D. H. & Froese, A. J. *Allergy clin. Immun.* **61**, 320–330 (1978).
4. Benveniste, J. *Clin. Allergy* **11**, 1–11 (1981).
5. Camussi, G., Tetta, C., Coda, R. & Benveniste, J. *Lab. Invest.* **44**, 241–251 (1981).
6. Pirotzky, E. *et al. Lancet* **i**, 358–361 (1982).
7. Yeung-Laiwah, A. C., Patel, K. R., Seenan, A. K., Galloway, E. & McCulloch, W. *Clin. Allergy* **14**, 571–579 (1984).
8. Reilly, D. T., Taylor, M. A., McSharry, C. & Aitchison, T. *Lancet* **ii**, 881–886 (1986).
9. Davenas, E., Poitevin, B. & Benveniste, J. *Eur. J. Pharmac.* **135**, 313–319 (1987).
10. Bastide, M., Doucet-Jacoeuf, M. & Daurat, V. *Immun. Today* **6**, 234 (1985).
11. Poitevin, B., Davenas, E. & Benveniste, J. *Br. J. clin. Pharmac.* **25**, 439–444 (1988).
12. Stanley, H. E., Teixeira, J., Geiger, A. & Blumberg, R. L. *Physics* **106A**, 260–277 (1981).
13. Fröhlich, H. *Adv. Electron. & Electron. Phys.* **53**, 85–152 (1980).
14. Smith, C. W. & Aarholt, E. *Hlth Phys.* **43**, 929–930 (1988).
15. Petiot, J. F., Sainte-Laudy, J. & Benveniste, J. *Ann. Biol. Clin.* **39**, 355–359 (1981).

## Editorial reservation

READERS of this article may share the incredulity of the many referees who have commented on several versions of it during the past several months. The essence of the result is that an aqueous solution of an antibody retains its ability to evoke a biological response even when diluted to such an extent that there is a negligible chance of there being a single molecule in any sample. There is no physical basis for such an activity. With the kind collaboration of Professor Benveniste, *Nature* has therefore arranged for independent investigators to observe repetitions of the experiments. A report of this investigation will appear shortly. □

# Metal enrichment in bauxites by deposition of chemically mature aeolian dust

George H. Brimhall\*, Christopher J. Lewis\*, Jay J. Ague\*, William E. Dietrich\*, Joachim Hampel\*, Tim Teague\* & Peter Rix†

\* Department of Geology and Geophysics, University of California at Berkeley, California 94720, USA

† Alcoa of Australia Limited, Applecross, Western Australia

*Mass-balance equations applied to chemical analyses of a bauxite weathering profile in Western Australia show quantitatively that the ore-grade enrichment of aluminium is due to accumulation of aeolian dust derived elsewhere from chemically mature soils. This finding challenges the prevalent view of bauxite genesis by simple in situ residual enrichment and suggests that regional dust trajectories may serve as a genetic link between laterites, bauxites, heavy mineral beach sands, and clay-enriched marine sediments.*

SINCE 1846, when Charles Darwin recognized the importance of atmospheric transport of land-derived dust to long-term offshore marine sedimentation<sup>1</sup>, there has been a growing awareness of the importance of dust deflation, transport and deposition over geological timescales<sup>2,3</sup>. Dust deposition has been implicated in many surficial processes including evolution of landscapes in desert marginal lands<sup>4</sup>, development of duricrusts such as calcrete layers by  $\text{CaCO}_3$  influx to form caliche<sup>5-7</sup>, loess formation<sup>8</sup> and significant transfer of elements from continents to the oceans<sup>9</sup>.

Ferruginous laterites and bauxites are widespread soils of considerable economic importance, serving both as mineable sources of Ni, Co, Al and Au and as nutrient-limited but intensively used agricultural lands<sup>5</sup>. For over a century, the origin of these soils has been debated, with arguments resting largely on stratigraphic, topographic and mineralogical observations. What is needed to resolve this debate, and to investigate the problem of continental weathering in general, is a means to evaluate the role in weathering profile development of dissolution, strain (volume decrease or increase), reprecipitation and vertical infiltration of detritus from external sources.

We have recently shown that a simple mass-balance approach can offer such quantitative information about the genesis of shallow supergene enrichment of Cu and Ni in laterites<sup>10,11</sup> and Al, Fe and Pb in podzols<sup>11</sup>. To test this approach in other weathering environments, we selected a well-studied, essentially 'classic' bauxite locality in the Darling Range of Western Australia<sup>12,13</sup> positioned along a well-known regional dust trajectory<sup>14</sup>.

The present dust trajectories in Australia extend both westward and eastward from the arid continental interior towards adjoining Indian and Pacific Oceans. On both sides of the continent, marine sediments are dramatically enriched in kaolinite within tongues extending thousands of km offshore<sup>14</sup> (Fig. 1a). In Western Australia the trajectory of easterly winds blowing offshore is aligned with the onshore desert sand-dune trend<sup>15</sup>. Although the major dust trajectory is towards the west, the wind pattern affords ample opportunity for transmigration, that is, back and forth movement. In summer, strong east to west winds carry windborne dust from the desert regions of the Yilgarn Craton to the coast. In winter, on the other hand, onshore westerly winds coming from the ocean to the southwest bring the bulk of present-day rainfall and salts, which are carried well into the interior basins of southwest Australia<sup>16</sup> (Fig. 1a).

Bauxite weathering profiles are particularly useful for dust deposition analysis as their surfaces are often covered by coarse pisolites and hence may be stable to wind erosion, though much

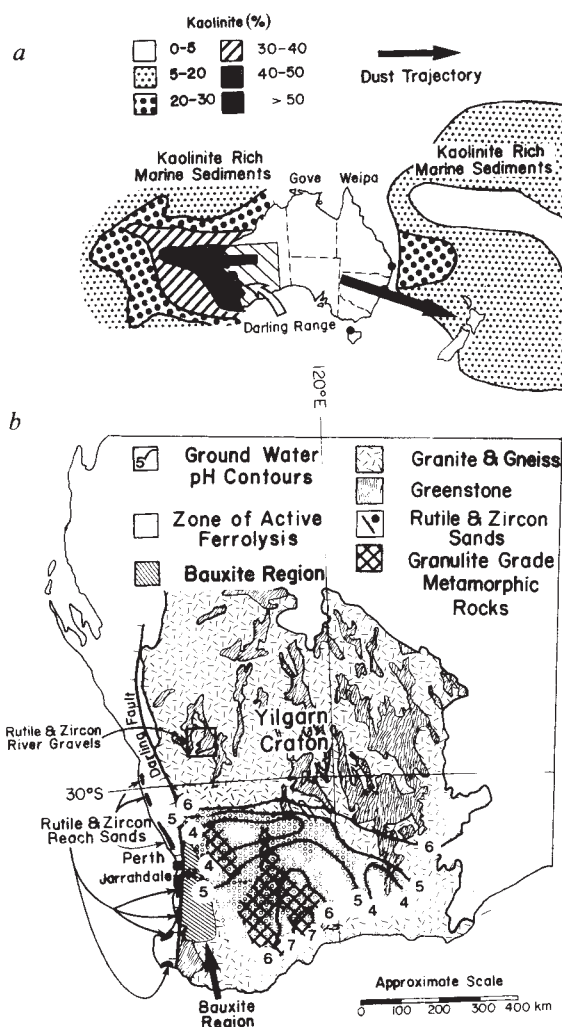
dust may blow across them with little net deflation<sup>2</sup>. The migratory dust path can be evaluated by determining the mineralogy of the weathering suite in each zone and ascertaining the presence of stable mineral detritus transported from external surficial deposits. Implementing this apparently simple strategy necessitates separation of chemical and mineralogical enrichments owing to several different processes: (1) residual enrichment of chemically inert species simply by *in situ* concentration during removal of more soluble mobile species, (2) enrichment due to deformation, that is volume change, and (3) actual introduction of exotic minerals, as dust from external sources, either nearby or distant. Analysis of these effects requires field description, petrographic (microscopic) examination, chemical analyses, and determination of physical properties such as density to implement, a relevant mass-balance model yielding chemical gains and losses in a well-characterized parent material<sup>10,11</sup>.

## Jarrahdale bauxite deposit

In applying our analytical techniques we have chosen a metasomatic system which, because of the homogeneity of the parent material, offers ideal opportunities for yielding unambiguous and accurate interpretation of surficial metasomatic processes. For this study, we report results from the Jarrahdale bauxite deposit in the Darling Range of Western Australia, which has developed on the Archaean Yilgarn craton consisting of gneisses, greenstones and granitic rocks (Fig. 1b). The weathering profile is pedogenically differentiated into five zones, which from bottom to top are: granodiorite parent material, saprolite, blocky hardcap, massive hardcap, and pisolitic gravel (Fig. 2). **Micromorphology of bauxite.** Petrographic examination of hardcap samples revealed a striking geopetal fabric (Fig. 3a, b). This presents clear evidence of the introduction of fine-grained mineral grains from above and subsequent grain sedimentation in voids. Crescentic microsedimentary layers consisting largely of gibbsite, kaolinite and iron oxides occur, each with a consistent concave upward orientation similar to clay structures resulting from translocation in argillic horizons of common soils<sup>17,18</sup>.

Continued pore-space evolution by progressive infilling and development of new voids are manifested in upward increase in the radius of curvature of microsedimentary structures within each void as well as by sedimentary unconformities produced by new void formation (Fig. 3a). Such features show that translocation of fine-grained mineral grains is a dominant feature of the fabric of the hardcaps. This being so, one must ask how much of the aluminium and iron enrichment so characteristic of bauxites in general is due to influx of ferruginous and aluminous minerals from external sources.



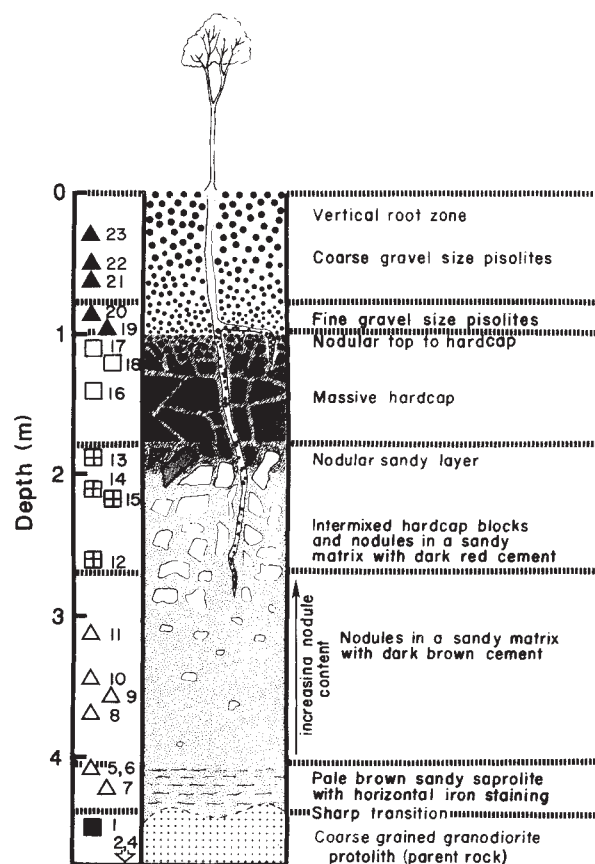


**Fig. 1** *a*, Westward and eastward dust trajectories extending from the arid interior of Australia to the Indian and Pacific Oceans (from Pye<sup>2</sup> and Windom<sup>14</sup>). Kaolinite-rich marine sediments extend along the dust trajectories. Shoreline heavy-mineral beach deposits are shown as black dots. *b*, Location of the Jarrahdale bauxite deposit in the Darling Range of Western Australia adjacent the Darling Range Fault. The sampling locality is the same deep railroad cutting as studied in detail by Sadleir and Gilkes<sup>21</sup> and compared with other world-class bauxite deposits elsewhere in northern Australia at Weipa and Gove<sup>13</sup>. Regions of active ferrolysis and surficial heavy-metal remobilization occur immediately east of the bauxites where acid saline lakes and groundwater are prevalent<sup>20</sup>. Granulite terrains are from Vallance<sup>21</sup>.

## Chemical and mechanical processes

A chemically immobile element must be identified with which to gauge the gains and losses of other, more mechanically and chemically mobile elements, and with which to calculate strain<sup>11</sup>. In these bauxites there are two possibilities: Zr and Ti. We find, however, that as well as being residually concentrated, both of these elements have themselves been introduced into the profile as zircon and rutile. As key stable minerals, the extent of Zr and Ti enrichment by introduction must be determined before introduction of other elements can be studied.

**Depth of zircon and rutile translocation.** We prepared concentrates of zircon and rutile from each sample by dissolving away more abundant minerals with a weak HF and aqua regia leach solution. By leaching samples without crushing, we retained the original size and morphology of these accessory minerals, which were resistant to acid attack. Rutile grains were invariably highly rounded and restricted to a near-surface zone extending from



**Fig. 2** Weathering profile in exposed railroad cutting through Craiges Ridge, 0.7 km southwest of the Alcoa no. 2 mine site. A total of 23 samples (1129-1 to 1129-23) were taken as shown. Symbols represent sample type: bedrock (■), saprolites (△), hardcap (□) and pisolites (▲). Quantitative modal petrography of the granite protolith (samples 1-4) was determined by computer-assisted line integration with a digitizing microscope<sup>19</sup>. The average mode of the granodiorite consists of 52% plagioclase, 32% quartz, 10% microcline and 3% biotite. The remaining 3% consists of the igneous accessory minerals, sphene, magnetite, apatite and zircon and alteration epidote, calcite and chlorite in decreasing order of abundance. No rutile or anatase was identified. The overlying saprolite becomes progressively richer in nodules and blocks of hardcap in its upper portions. Above the hardcaps, pisolites become progressively coarser upwards. Bedrock and hardcap were sampled by hammer and chisel while soils were samples with a hand-held piston coring device from which undistributed soil cores 48 mm in diameter and 77 mm in length were taken and recovered inside durable plastic tubes. During soil sampling no compression of the soils was observed, so that subsequent density measurements are representative of the in-place soils.

the pisolites down through the blocky hardcap (Fig. 4a). No rutile was present in the parent granodiorite. The mean diameter, maximum diameter and abundance of rutile grains decreased consistently with depth as determined with the digitizing microscope<sup>19</sup>.

Two populations of zircon occur: colourless, well-rounded grains occurring only in the pisolitic zone, and cloudy, perfectly euhedral and zoned grains ubiquitous in the profile and clearly derived from weathering of the granodiorite. The average zircon grain diameter decreased with depth to the base of the hardcap, but then abruptly increased in the upper saprolite to a stable value similar to that of the granodiorite (Fig. 4b). We interpret the rutile and zircon grain size, morphology and abundance profiles above the base of the hardcap as resulting principally from translocation of rounded rutile and zircon downward from the surface.

**Pore size restrictions: a porous media sieve.** By determining the average pore diameter in stained epoxy-impregnated thin sections we have shown that pore diameter decreases with depth (Fig. 4c). The depth of translocation is therefore governed by the relative sizes of translocated grains and the diameter, shape and tortuosity of pore structures encountered during passage through the bauxite. We conclude that rutile and zircon translocation is limited by pore size, which in turn is controlled by the extent of pore infilling by Al and Fe compounds and continued void formation by eucalyptus tree roots. The bauxite profile has clearly behaved as a porous media sieve that tends to trap infiltrating exotic and neoformed mineral grains, producing a grain-size sorting effect.

### Separation of fractionation processes

Here, *p* stands for properties of the parent material or protolith, residual increase in concentration of chemically immobile elements, (2) internal collapse or strain, and (3) translocation of exotic grains can each be quantitatively assessed<sup>10,11</sup>. Zr is determined by X-ray fluorescence and other metals investigated were determined by flame or flameless atomic absorption spectroscopy.

**Residual enrichment.** Using constitutive mass-balance equations relating chemical concentration to density and volume of weathering zones, we have demonstrated previously<sup>10,11</sup> the utility of expressing metasomatic changes in terms of the ratio of the concentration of a given element (*i*) in a weathered sample ( $C_{i,w}$ ) to the concentration in the parent material ( $C_{i,p}$ ). This unitless fractionation factor,  $C_{i,w}/C_{i,p}$ , is greater than unity for enrichment, or less than unity for depletion of an element, *i*. These fractionation factors for the two most important elements in bauxite, Fe and Al, are shown in Fig. 5. We have shown that in the absence of deformation and translocation effects, the enrichment factor due to simple residual concentration is a simple function of density<sup>11</sup>:

$$C_{i,w}/C_{i,p} = \rho_p/\rho_w \quad (1)$$

Here, *p* stands for properties of the parent material or protolith, *w* for properties of the weathered material, and  $\rho$  is bulk density. Figure 5 shows the overall enrichment factor for each sample with a symbol for each depth zone. The component of the overall enrichment factor due to simple residual enrichment, given by equation (1), is shown with a dashed line and accounts for only a small fraction of the overall Fe and Al concentration, indicating that translocation and strain must be important.

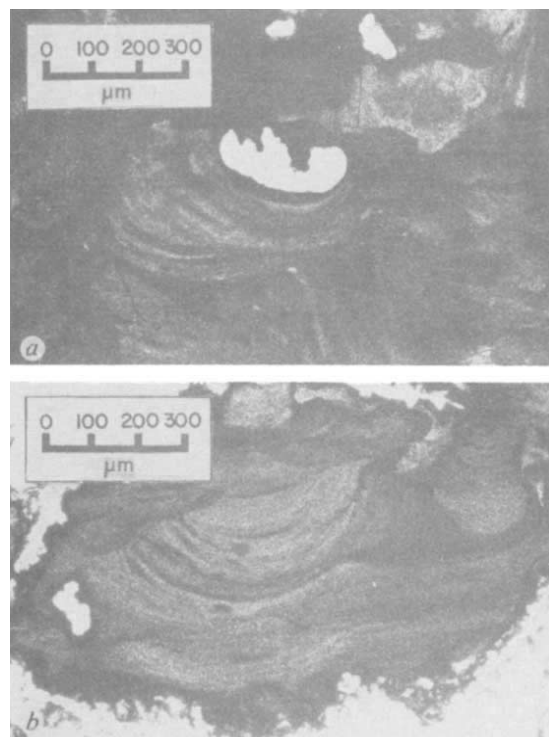
**Volumetric change (strain).** Strain due to weathering,  $\varepsilon_{i,w}$ , is defined here in one dimension as the change in length divided by the original length. If the vertical dimension of a representative volume of protolith before weathering is  $B_p$  and after weathering is  $B_w$ , the strain  $\varepsilon_{i,w}$  is simply  $(B_{i,w} - B_{i,p})/B_{i,p}$ . In addition it can be shown<sup>11</sup> that strain can be related to the fractionation factor:

$$\frac{C_{i,w}}{C_{i,p}} = \frac{\rho_p}{\rho_w} \left( \frac{1}{\varepsilon_{i,w} + 1} \right) \quad (2)$$

We determine the amount of chemical enrichment due to strain by first applying our measurements on the depth of heavy mineral translocation. Below this depth, any enrichment of Ti and Zr in excess of residual enrichment is due to deformation, as no evidence of dissolution or reprecipitation of rutile or zircon exists.

The net strain due to weathering,  $\varepsilon_{i,w}$ , is determined from equation (2), which is a complete closed-system form of equation (1).

**Combined residual enrichment and strain.** Figure 5 shows the component of the overall enrichment factor (data points) accounted for by strain and residual enrichment, using a solid line based on equation (2). This function approximates the whole-rock data much more closely than residual enrichment alone (dashed line). But near the top of the profile, even this



**Fig. 3** Micromorphological features of blocky hardcap sample (1129-12) shown in plane polarized light. All massive and blocky hardcap samples of the bauxite profile were impregnated with blue-coloured epoxy under vacuum in order to prepare thin (30  $\mu\text{m}$ ) and ultra-thin (5  $\mu\text{m}$ ) sections so that delicate pore structures were preserved and could be measured. Partly (a) and completely (b) infilled pores with crescentic microclastic deposits (films) of gibbsite, kaolinite and iron oxides filling voids within blocky hardcap (sample 1129-12). The radius of curvature of geopetal infillings increases as pores decrease in size. A micro-sedimentary unconformity is apparent in the lower right-hand side of (a).

combined function accounts for only a small part of the overall enrichment; the remainder is due to open-system enrichment. **Open-system transport ( $\tau$ ).** To measure the addition or extraction of material either by translocation or solute migration, equation (2) can be expanded to

$$\frac{C_{i,w}}{C_{i,p}} = \frac{\rho_p}{\rho_w} \left( \frac{1}{\varepsilon_{i,w} + 1} \right) + \frac{100 m_{i,\text{influx}}}{C_{i,p} \rho_w B_w} \quad (3)$$

where  $m_{i,\text{influx}}$  is the mass of element *i*, added to an elementary volume of vertical dimension  $B_w$ .

**Translocation.** Solving equation (3) for the mass influx,  $m_{i,\text{influx}}$ , and dividing by the mass of element *i* contained initially in the original parent interval,  $C_{i,p} \rho_p B_p$ , yields

$$\tau_{i,w} = \frac{100 m_{i,\text{influx}}}{C_{i,p} \rho_p B_p} = \frac{\rho_w}{\rho_p} \frac{C_{i,w}}{C_{i,p}} (\varepsilon_{i,w} + 1) - 1 \quad (4)$$

where  $\tau_{i,w}$  is mass added to the unweathered closed system, which is zero when there has been no translocation. If  $\tau$  is positive, an influx of material has occurred: if  $\tau$  has a value of +5, for example, translocation accounts for five times or 500% the mass of element *i*, contained in the unweathered material. If  $\tau$  is negative, material has left the system in suspension. The minimum possible value of  $\tau$  is -1.0 or -100% for total depletion. The utility of the transport function,  $\tau$ , is in recognizing regions within the weathering profiles where elements have been



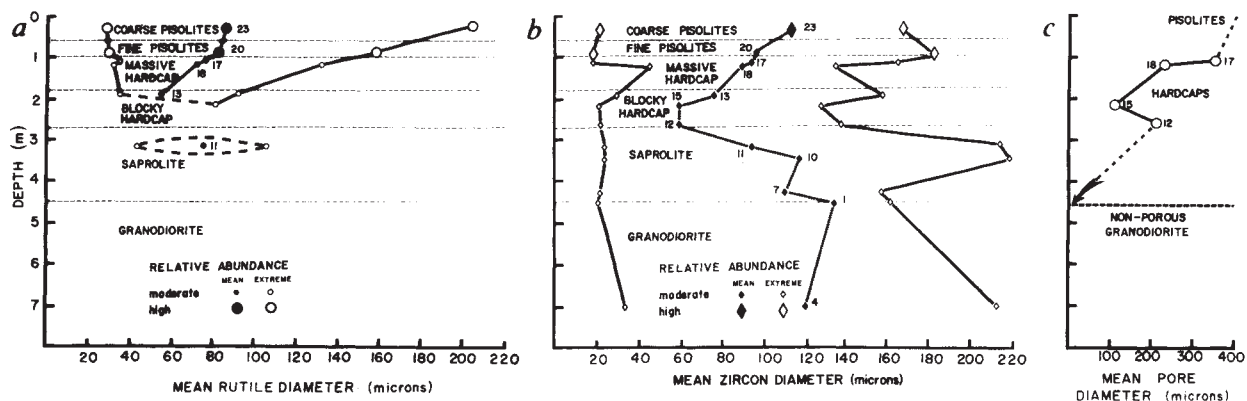


Fig. 4 a, Distribution, relative abundance and size range of rutile with depth in the weathering profile. The average and maximum grain size of rutile decreases with depth. Rutile is present only above the saprolite and is a transported, exotic mineral. b, Distribution, abundance and size range of zircon. The average grain size of zircon decreases to a depth of 2.5 m, below which the fine-grained, rounded and abraided exotic variety no longer persists. c, Average diameter of pores in the hardcaps. There is an overall decrease from very large macropores in the pisolitic zone to an absence of even  $\mu\text{m}$ -signal pores in the granodiorite bedrock.

added or removed (Fig. 5). The pore fillings in the bauxite hardcaps suggest that mass influx occurs by translocation of mineral particles rather than reprecipitation of solutes in solution, although microchemical precipitates occur as void infillings (Fig. 3a). Near the top of the profile both Fe and Al have been added to the system, whereas near the saprolite-bedrock interface small amounts of these elements have been removed. One of the highest observed values of  $\tau$  is  $\sim 1,000\%$  for Fe, indicating that for each gram of Fe in the parent material 10 g has accumulated. For Al,  $\tau$  approaches 300%, showing that residual enrichment accounts for only a very small fraction of the Al in the profile. Most is added by translocation by Al-rich detrital minerals formed by chemical weathering elsewhere. But as the new transport function,  $\tau$ , is not indicative of the distance of transport before influx of grains into the profile, the material could be entirely allochthonous or, if derived more locally, parautochthonous, or both. Figure 6 shows that a suite of other elements have similarly been systematically introduced into the top of the profile. This suite included Zr, Ti, V, Cr and Mo. None of these elements have local source regions ( $\tau = 0$ ) within the profile.

### External source of heavy metals

The particular suite of introduced metals (Fe, Al, Zr, Ti, Cr and Mo) in combination with the lack of any evidence of quartz enrichment ( $\tau$  for Si is slightly negative, consistent with retention of primary quartz) is indicative of a source material peculiarly enriched in these elements and surprisingly deficient in quartz. This enigma is resolved by considering specific surficial processes and bedrocks in the region immediately east (upwind) of the bauxite studied (Fig. 1b). Here hydrogeochemical zoning patterns in the Yilgarn block have developed by shallow redistribution of elements during weathering, including concentration of Fe, Pb, U and Au<sup>20</sup>.

Saline seepages, salt lakes and playa systems span  $>80\%$  of this region. High concentrations of Fe and Al have been recognized in acid saline groundwaters with  $\text{pH} < 4$  through ferrololysis, a process by which  $\text{Fe}^{2+}$  is oxidized to  $\text{Fe}^{3+}$  above the groundwater table and precipitated as iron oxide with the liberation of  $\text{H}^+$  (ref. 20):



The region of active ferrololysis is characterized by heavy-metal remobilization, particularly those metals whose solution chemistry is controlled by a change in oxidation state<sup>20</sup>. We hypothesize that the regional source of the transported heavy-metal suite contaminating the upper portions of Jarrahdale bauxite profile are surficial laterite and ferricretes, deficient in

quartz but enriched in rutile and zircon and metals due to intense ferrololysis in Western Australia. The occurrence of granulite grade<sup>21</sup> metamorphic rocks (Fig. 1b) containing orthopyroxene, which is exceptionally susceptible to rapid low-temperature oxidation to haematite and kaolinite<sup>22</sup>, suggests that regional lithological factors may also contribute to ferrololysis (R. C. Newton, personal communication). The metallic elements, concentrated near the surface, may be transported as complex dust and detrital primary minerals in the aeolian dust. Elsewhere, for example in Nigeria, iron-oxide-rich dust blown southward from the Sahara Desert has been implicated as a major additive in soils<sup>23</sup>. Although aluminous weathering products such as gibbsite and kaolinite are unquestionably developed *in situ* by hydrolysis reactions of feldspar, and chemically resistant heavy minerals such as zircon are residually enriched as well, major addition of detrital phases has occurred and dominates the chemistry of the top the Jarrahdale bauxite profile. The high porosity and pore structure of bauxites are ideal for translocation and are developed through continued pedoturbation by tree

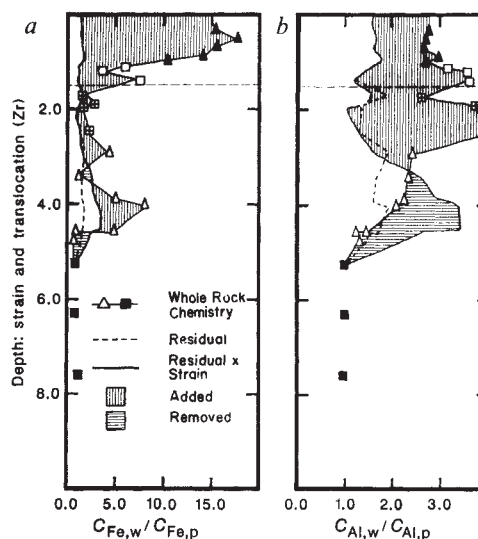
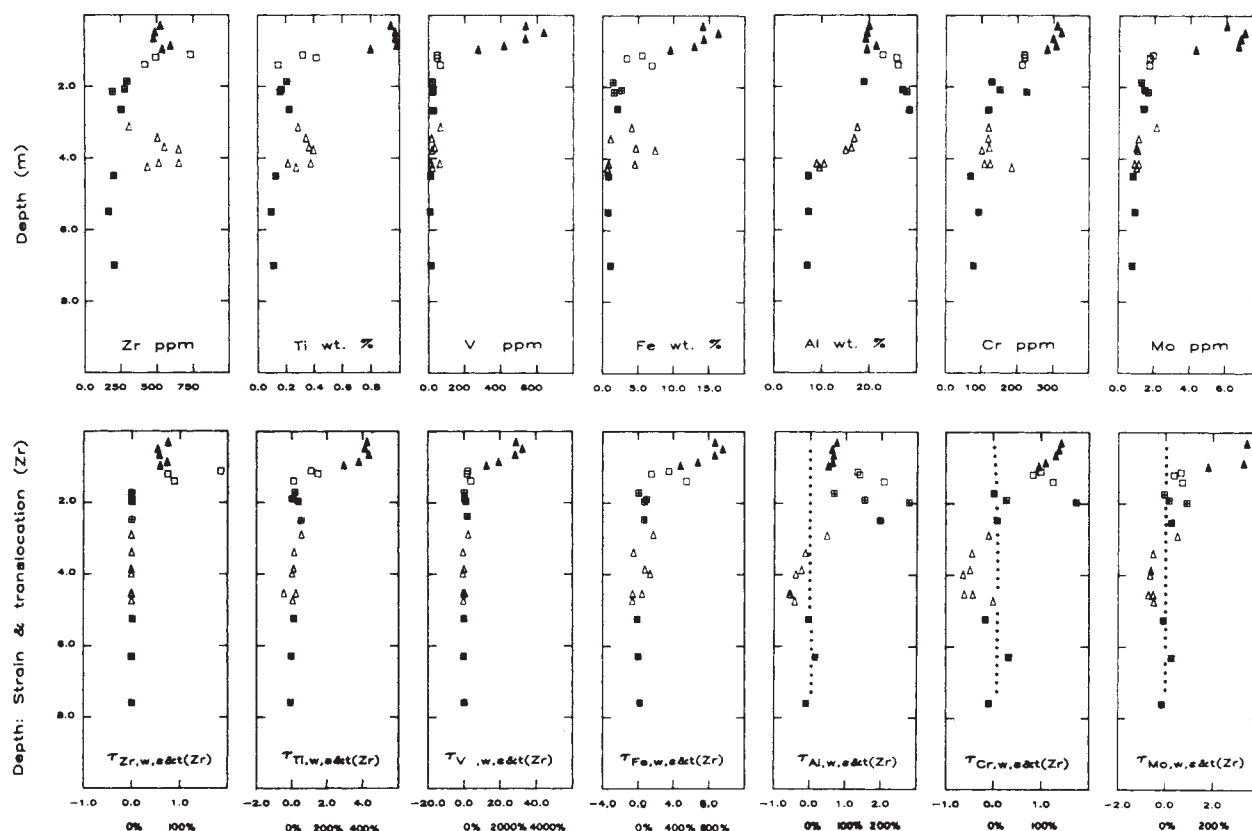


Fig. 5 Contributions of residual enrichment, strain and translocation to enrichment of Fe (a) and Al (b) using Zr for strain determination. Symbols indicate the horizons shown in Fig. 2. Residual enrichment, represented by the dashed line, is simple  $\rho_{\text{Fe}}/\rho_{\text{w}}$ . Most of the Fe and Al is introduced into the profile from above. Depths shown are corrected for collapse. Zones where Zr translocation has been demonstrated are assumed to have experienced no strain.



**Fig. 6** Variation in mechanical transport function,  $\tau$ , with undeformed depth for Zr, Ti, V, Fe, Al, Cr and Mo, using Zr to calculate strain. Effects of residual and deformational enrichment are taken into account. Vertical dotted line indicates zero translocation. Positive values of  $\tau$  indicate addition to the profile; negative values, subtraction. All of these elements have been added to the top of the profile. Al, Cr and Mo have also been slightly removed from the bottom of the profile.

roots forming a network of channels and tubular voids which provide access for surficial grains to the subsurface<sup>24</sup>. The replacement of lithostructure by pedostructure occurs through infilling of voids with exotic sediment.

**Dust deposition.** During weathering, continued partial erosion of bauxite and laterite profiles supplies a pre-enriched material for subsequent transport and recycling of these stable and resistant mineral species to adjacent regions to the west. The resultant chemical enrichment is in fact largely mechanical at the site of deposition. Besides weak residual and volumetric enrichment, fractionation occurs by translocation and porous media sieving, as grain-size sorting is affected by progressive impedance of grains by pore restrictions. The exotic material added to bauxites are mature detrital grains that have been released from their previous surficial position. Hence they are largely unreactive within their new environment and may persist indefinitely without modification until later released by abrasion or erosion. We therefore view bauxite genesis as a continuing cumulative fractionation process involving dust migration, with eclectic grain mixtures developing along the dust trajectory by temporary storage in the shallow subsurface. Although cumulative soil formation involving addition of detrital parent material during weathering is not a new idea<sup>18,24,25</sup>, for instance allochthonous karst bauxite development on limestones in the Mediterranean<sup>26,27</sup>, this is the first conclusive proof that addition of aeolian material perhaps in combination with parautochthonous enrichment products can be a major factor in the genesis of true bauxites occurring along a regional dust trajectory.

**Mineralogical anomalies.** With this conclusion we can reconcile mineralogical anomalies previously recognized by Grubb<sup>28</sup> with the simple residual bauxitization of granitic bedrock proposed earlier<sup>24,29-33</sup> by *in situ* transformation involving extreme fractionation by chemical selectivity<sup>34</sup>. Within the upper portions

of numerous profiles in the Darling Range a distinctive, abraded heavy-mineral suite consisting of zircon, rutile and ilmenite, each of several varieties, was interpreted by Grubb<sup>28</sup> to be largely exotic in origin as many laterites contained components not found in the bedrock. Grubb attributed this detrital suite to residual weathering of a composite parent material consisting of a coarse feldspathic fluvialite mantle of sediments inferred to have once overlain the Archaean granite bedrock. Sadleir and Gilkes<sup>12</sup> criticized this hypothesis, and demonstrated that bauxite mineralogy varied with underlying bedrock lithology of gneisses and basaltic dikes. They also cited the absence of any fresh rock at the base of the profiles Grubb<sup>28</sup> studied as being another deficiency in his interpretation. They consequently concluded that the bauxite was a residual product of weathering, but they failed to resolve the origin of the abraded heavy-mineral suite. Glassford and Killigrew<sup>35</sup> argued that desert conditions once existed throughout the southwest of Australia. Davy<sup>36</sup> concluded subsequently that windblown minerals would therefore be expected over this expansive region, but stated that the origins of the exotic accessory minerals and the means of their incorporation into the caprock had not been clearly identified. We show that the only weathering process consistent with interpretation of the new chemical, physical and microtextural data is that, rather than derivation from the inferred sedimentary mantle for which there is little geological evidence, the abraded mineral suite and, surprisingly, the major portion of the Al and Fe were introduced into the top of the bauxite profiles by widespread aeolian activity (Fig. 1a). Unlike local residual enrichment, which we show to be a minor factor, translocation of minerals of either allochthonous or parautochthonous origin can produce large detrital enrichment factors not only for accessory minerals but also for major ore-forming minerals such as kaolinite and gibbsite.



**Association of bauxites and heavy-mineral beach sands.** As kaolinite is only one of the characteristic secondary minerals enriched by intense chemical weathering, chemically resistant primary mineral grains concentrated by chemical weathering must similarly be transported westwards from the continental interior, though not necessarily the same distance as kaolinite. If the entire mixed-weathering mineral suite consisting of stable, chemically mature minerals becomes airborne, then the denser, coarser detritus may be deposited onshore in a closer subaerial environment than the offshore accumulation of kaolinite (Fig. 1a). We base this inference on the surficial deposits observed along the dust trajectory, each of which contains both primary and secondary mineral grains that are relatively stable with respect to chemical weathering. The east-west sequence (Fig. 1b) consists of laterite and ferricrete duricrust regions within closed acid saline basins<sup>20,37-39</sup>, nearby proximal bauxite deposits in the Darling Range<sup>12,13</sup>, shoreline beach deposits enriched in detrital heavy minerals including zircon and rutile<sup>40-42</sup>, and finally distal offshore submarine pelagic sediments enriched in kaolinite<sup>14</sup>. We postulate that the surficial deposits in this sequence are genetically related and result from progressive mechanical fractionation and selective deposition of the aeolian mineral population. Because of their high porosity

the upper portions of bauxite and laterite weathering profiles along the trajectory may be temporary, though long-lived, repositories for airborne dust derived from source regions toward the east, therefore offering practical means of sampling long-term dust deposits. This conclusion leads us to look globally at the distribution of large bauxite and laterite deposits occurring at tectonically stable continental margins (Australia, Ghana, Guinea and Sierra Leone in West Africa and India) where world-class heavy-mineral (rutile, zircon, monazite) shoreline deposits occur<sup>41-44</sup>, and to infer that major offshore dust trajectories, in combination with fluvial systems, control migration paths and fractionation of mature mineral suites on their long-term passage to the sea. Efficient cumulative enrichment involving continental weathering occurs by a combination of stripping apparently older surficial debris, its aeolian transport, and entrapment in more humid coastal environments after derivation from an almost infinite source region undergoing aridification.

We thank Roger Vines and Shane Sadleir for assistance with the field work, which was coordinated by Roy Woodall; we also thank Oliver Chadwick, Graham Reynolds, Peter Ball, Charles Gilbert, Hans Jenny, Kathy Danti and Shari Shute for discussions and assistance.

Received 21 December 1987; accepted 22 April 1988.

1. Darwin, C. Q. *geol. Soc. Lond.* **2**, 26-30 (1846).
2. Pye, K. *Aeolian Dust and Dust Deposits* (Academic, London, 1987).
3. Pewe, T. L., Pewe, E. A., Journaux, A. & Slatt, R. M. *Spec. Pap. geol. Soc. Am.* **186**, 169-190 (1981).
4. McTainsh, G. H. & Pitblado, J. R. *Earth Surf. Process Landforms* **12**, 415-424 (1987).
5. Coque, R. *Ann. Geol.* **64**, 359-370 (1955).
6. Goudie, A. S. *Duricrusts* (Clarendon, Oxford, 1973).
7. Watson, A. *Sedimentology* **32**, 855-876 (1985).
8. Richthofen, F. von *Geol. Mag.* **9**, 293-305 (1882).
9. Goldberg, E. D. *Comments Earth Sci. Geophys.* **1**, 117-196 (1971).
10. Brimhall, G. H., Alpers, C. & Cunningham, A. B. *Econ. Geol.* **80**, 1227-1254 (1985).
11. Brimhall, G. H. & Dietrich, W. A. *Geochim. cosmochim. Acta* **51**, 567-587 (1987).
12. Sadleir, S. B. & Gilkes, R. J. *J. geol. Soc. Aust.* **23**, 333-344 (1976).
13. Loughnan, F. C. & Sadleir, S. B. in *Bauxite* (ed. Jacob, L.) 436-450 (Am. Inst. Metallurg. Petrol. Engrs, New York, 1984).
14. Windom, H. L. *J. Sedim. Petr.* **45**, 520-529 (1975).
15. Brookfield, M. Z. *Geomorph. Suppl.* **10**, 121-158 (1970).
16. Hingston, F. & Gailitis, V. *Aust. J. Soil Res.* **14**, 319-335 (1976).
17. McKeague, J. A. in *Soil Micromorphology* (eds Bullock, P. & Murphy, C. P.) 369-371 (Academic, Berkeley, 1981).
18. Birkeland, P. W. *Pedology, Weathering and Geomorphological Research*, 146-147 (Oxford University Press, New York, 1974).
19. Brimhall, G. H. & Rivers, M. L. *United States Patent* **4**, 503, 555 (1985).
20. Mann, A. W. in *Geochemical Exploration in Deeply Weathered Terrain* (ed. Smith, R. E.) 158-167 (CSIRO Inst. Energy and Earth Resources, Wembley, 1983).

21. Vallance, T. G. *Metamorphic Map of Australia* (Bureau of Mineral Resources, Sydney, 1983).
22. Janardhan, A. S., Newton, R. C. & Hansen, E. C. *Contr. Miner. Petr.* **79**, 130-149 (1982).
23. Vine, H. *Tech. Commun. Commonwealth Bur. Soil Sci.* **46**, 22-29 (1949).
24. Tardy, Y. & Nahon, D. *Am. J. Sci.* **285**, 865-903 (1985).
25. Nikiforoff, C. C. *Soil Sci.* **67**, 219-230 (1949).
26. Vadász, E. *Bauxitfeldtan* (Akadémiai Kiadó, Budapest, 1951).
27. Valetton, I. *Bauxites*, 145-147 (Elsevier, Amsterdam, 1972).
28. Grubb, P. L. C. *Econ. Geol.* **66**, 1005-1016 (1971).
29. Baker, G. F. U. *Bull. Bur. Miner. Resour. Geol. Geophys. Aust.* no. 141, 223-224 (1973).
30. Geidans, L. *Australas. Inst. Min. Metall. Conf. Perth* **3**, 173-182 (1973).
31. Hutton, J. T. *Div. Rep. Div. Soils CSIRO* **11**, 55-56 (1955).
32. Tomich, S. A. *Proc. Australas. Inst. Min. Metall.* **212**, 125-135 (1965).
33. Valetton, I. in *Lateritisation Processes* (Proc. Int. Seminar on Lateritisation Processes) 15-23 (Balkem, Rotterdam, 1981).
34. Norton, S. A. *Econ. Geol.* **68**, 353-361 (1973).
35. Glassford, D. K. & Killigrew, L. P. *Search* **7**, 394-396 (1976).
36. Davy, R. *Geol. Surv. Western Aust.*, Report 8 (1979).
37. Mann, A. W. *Econ. Geol.* **79**, 38-49 (1984).
38. Mann, A. W. & Ollier, C. D. *Catena Suppl.* **6**, 152-157 (1985).
39. Webster, J. G. & Mann, A. W. *J. Geochem. Explor.* **22**, 21-42 (1984).
40. Fantel, R. J., Buckingham, D. A. & Sullivan, D. E. *Inform. Circ. U.S. Bur. Mines* **9061**.
41. Baxter, J. L. *West Austr. Geol. Surv. Min. Resour. Bull.* **10** (1977).
42. Baxter, J. L. in *Australia: A World Source of Ilmenite, Rutile, Monazite and Zircon*, 1-8 (Australian Institute of Mineral Metallurgy, Victoria, 1986).
43. Derry, D. R. *World Atlas of Geology and Mineral Deposits* (Mining Journal Books, London, 1980).
44. Dixon, C. J. *Atlas of Economic Mineral Deposits* (Cornell, Press, New York, 1979).

# DNA sequence determinants of CAP-induced bending and protein binding affinity

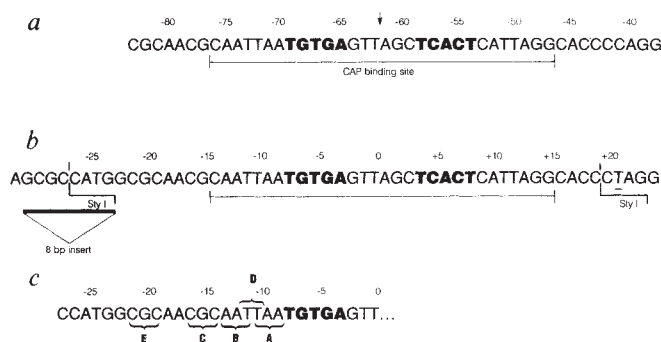
Marc R. Gartenberg and Donald M. Crothers

Department of Chemistry, Yale University, New Haven, Connecticut 06511, USA

*The sites of DNA bending induced by binding catabolite activator protein are identified and shown to coincide with positions where DNA grooves face the protein. The bendability of DNA with different sequences at these bend centres parallels the bending preference of the sequences in nucleosomal DNA. Anisotropic DNA bendability significantly affects the structure and strength of regulatory protein-DNA complexes.*

ELECTROPHORETIC analysis has revealed that the catabolite activator protein (CAP) bends its DNA binding site in the *Escherichia coli lac* promoter<sup>1-3</sup>. The accelerated cyclization kinetics of CAP-bound fragments<sup>4,5</sup>, hydrodynamic properties of CAP-DNA complexes<sup>6</sup> and models based on the protein's crystal structure<sup>7,8</sup> support this interpretation. Now protein-induced DNA bending is recognized as a structural motif common to many protein-DNA complexes<sup>9-17</sup>. The structural

requirements governing protein-induced bends are not fully understood; however, electrophoretic characterization of several mutant *lac* promoter-CAP complexes has demonstrated that DNA sequence determinants within the protein binding site modulate the extent of protein-induced bending. The L8 consensus mutation<sup>1</sup>, the right-hand symmetrized binding site<sup>18</sup>, and sequence changes outside the central 28 base pairs (bp)<sup>18</sup> reduce the overall bend.



**Fig. 1** *a*, Sequence and features of the wild-type CAP binding site in the *lac* promoter. A bracket denotes the thermodynamically defined binding domain<sup>18</sup>, bold face indicates nucleotides in the consensus region<sup>22</sup>, and an arrow identifies the pseudo-dyad axis of symmetry. Positions are numbered relative to the transcription start site. *b*, Changes required to introduce Styl restriction sites. Dinucleotide steps are numbered relative to the pseudo-dyad axis. Eight bp were inserted at step -23 (between positions -83 and -84) and the cytosine intersecting steps +20 and +21 (position -41) was replaced with thymine. A 211 bp *lac* promoter fragment containing this sequence was cloned into the EcoRI site of pGEM-2 (Promega). In this fragment, the CAP binding site is located ~80 bp from the upstream end. *c*, Location of sites fully randomized by oligonucleotide replacement. Each site was mutagenized independently of the others. Methods used were similar to those found in ref. 23.

**Table 1** Representative list of mutants and their attributes

| Site | Name       | Sequence             | $\mu_{rel}$ | $K_{rel}$ | $\sim\Delta$<br>(bending angle) |
|------|------------|----------------------|-------------|-----------|---------------------------------|
| A    | 11         | <sup>-11</sup> TTATT | 0.94        | 0.93      | 5°                              |
| A    | 24         | TAGCT                | 1.14        | 0.08      | -15°                            |
| A    | sdt        | TGCAT                | 1.29        | 0.07      | -30°                            |
| B    | 224        | <sup>-11</sup> CGAAT | 0.99        | —         | 0°                              |
| B    | 47         | CAAGT                | 1.08        | —         | -9°                             |
| B    | 206        | CAGCT                | 1.18        | —         | -19°                            |
| C    | 7          | <sup>-16</sup> ACCCA | 0.99        | 0.91      | 0°                              |
| C    | 99         | AGATA                | 1.08        | 0.85      | -9°                             |
| C    | 13         | ATTTA                | 1.16        | 0.79      | -17°                            |
| D    | 6          | <sup>-11</sup> AAAA  | 0.98        | —         | 1°                              |
| D    | 5          | ACCA                 | 1.26        | —         | -27°                            |
| E    | 25         | <sup>-21</sup> GATAA | 0.98        | —         | 1°                              |
| E    | 3          | GGCGA                | 1.00        | —         | 0°                              |
|      | wt         |                      | 1.00        | 1.00      | 0°                              |
|      | wt-129 bp  |                      | —           | 1.00      | —                               |
|      | syml       |                      | 0.86        | 24.6      | 13°                             |
|      | cons       |                      | 0.90        | 17.1      | 9°                              |
|      | symr       |                      | 1.31        | 0.04      | -32°                            |
|      | A22-218 bp |                      | 0.78        | —         | 20°                             |
|      | A27-219 bp |                      | 1.18        | —         | -20°                            |

Syml and symr were generated through symmetrization of dinucleotide steps 4-15 to the left and right of the binding site, respectively. A single T → A transversion at position -54 generates a mutant, cons, with a full consensus sequence. A22 and A27 result from the fusion of a wt binding site with a single in-phase and out-of-phase A tract, respectively. The appropriate spacing (22 bp) for the two bending elements in A22 was determined experimentally by methods similar to those found in ref. 26. It was assumed that an additional 5 bp would generate the out-of-phase construct, A27. 211-bp fragments were used unless otherwise specified. The mobilities of all mutant complexes were measured relative to those of wt complexes containing DNA fragments of equivalent length. Binding and bending assays were performed according to conditions specified in the caption to Fig. 2. Supercoiled plasmid sequencing methodology was adapted from the Promega Biotec manual (1986).

ized binding sites (see Fig. 1b)<sup>18</sup>. Synthetic oligonucleotides containing three successive base pairs of randomized sequence replaced the natural site. This procedure, referred to as cassette mutagenesis<sup>29</sup>, co-generated mutants that were resolved by bacterial transformation. All five independently randomized sites were outside and to the left of the consensus binding domain (see Fig. 1c). It is likely that mutations related by symmetry through the dyad axis would have a similar influence on structure. Using this approach, we isolated over 200 unique binding site mutants (see Table 1 for representative examples).

Mutant complexes were characterized by their electrophoretic mobility in non-denaturing polyacrylamide gels. According to the reptation theory of gel electrophoresis, DNA fragment mobility depends on the mean-square end-to-end distance. bent molecules snake through gel pores with more difficulty<sup>30,31</sup>. We have assumed that binding site mutations alter the extent of protein-induced DNA bending, but that the contribution of the protein to the electrophoretic mobility remains unchanged<sup>3,32</sup>.

## Correlation of mobility with sequence

The mobilities of mutant complexes vary from the value for the wt complex by as much as 30% (+30, -10) (see Fig. 2, lanes A-D). Few mutations in the non-consensus domain reduce the mobility of the complex, indicating that this region of the *lac* promoter has been optimized for near maximal bending. From

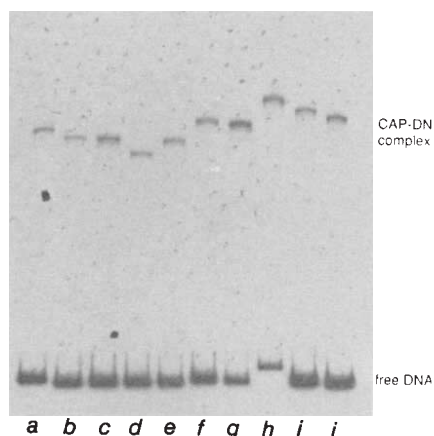
Properties of other protein-bound DNAs support the role of DNA sequence-dependent elasticity in modulating protein binding affinity. Statistical sequencing and Fourier analysis detect DNA sequence periodicities responsible for the rotational positioning of nucleosomal DNA on the histone octamer<sup>19,20</sup>. An analysis of 434 repressor binding affinity to mutant operators suggests that DNA sequence-dependent stiffness contributes to the specificity of that interaction<sup>21,22</sup>. The *lac* promoter CAP binding site differs in structure from these two examples. Unlike nucleosomal DNA, the site contains a consensus binding domain consisting of two symmetrically related (but separated) 5-bp segments<sup>23</sup> which interact specifically with the protein's helix-turn-helix recognition element<sup>7,8,24-26</sup> (see Fig. 1a). Unlike the 434 repressor binding site, the consensus is nested between adjacent distal binding domains which are required for full binding affinity<sup>18</sup>. The observed ramp of positive electrostatic potential on the protein surface<sup>8</sup> and the finding of little sequence homology with other CAP binding sites<sup>23</sup> implies that interactions in these distal binding domains are electrostatic, non-specific, and therefore primarily DNA structure-dependent.

The primary issues we address here are whether the sequence dependence of DNA bendability that has been deduced from the properties of nucleosomal DNA<sup>19,20,27</sup> applies also to regulatory protein-induced DNA bends (as proposed in ref. 28) and whether the variation in DNA bend energy is large enough to affect complex structure and binding affinity significantly. To this end, we have generated an extensive library of CAP binding site mutants and measured their effects on the electrophoretic mobility of CAP-DNA complexes in non-denaturing polyacrylamide gels. Competitive binding affinities and changes in bending angle were measured in selected cases. Our results strongly support the view that sequence-dependent anisotropy of DNA bendability is a dominant factor in modulating the contribution of the distal binding domains to the overall bend and binding affinity of the CAP-DNA complex.

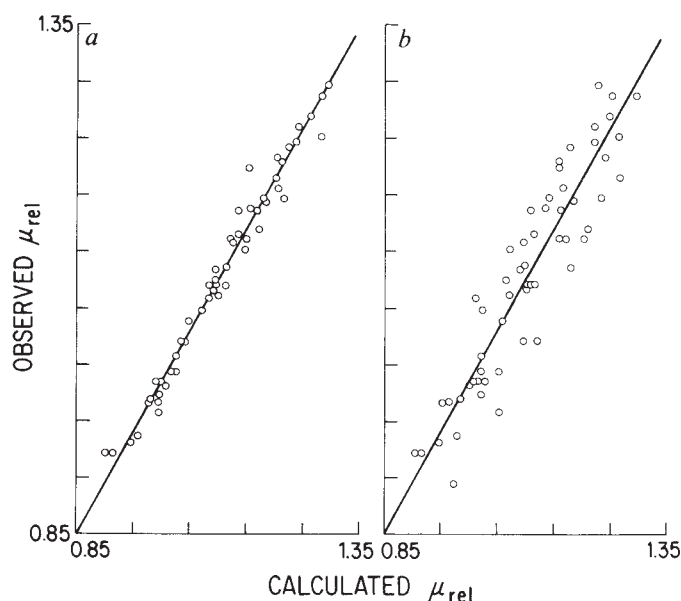
## Binding site construction

The incorporation of unique restriction sites flanking the CAP binding site in a *lac* promoter fragment allowed the removal and replacement of the wild-type (wt) sequences with mutagen-





**Fig. 2** CAP-DNA complexes with variant mobility. Lanes A-J are mutants 99Cx, 13Cx, 24Ax, sdtAx, A27, wtx, wty, A22, symly, and consy, respectively. The 'y' and 'x' suffixes indicate that these DNA fragments differ in length from those in Table 1 by an additional 6 and 7 bp at the upstream Styl site respectively. Though the lengths of the mutant and wt fragments were increased, relative mobilities did not change. Mutant *lac* promoter fragments were cleaved from parent plasmids with *Eco*RI (New England Biolabs), labelled with  $\alpha$ - $^{32}$ P-ATP (Amersham) and Klenow fragment (New England Biolabs), and isolated by polyacrylamide gel electrophoresis. CAP complexes were formed in 20- $\mu$ l reactions containing 100  $\mu$ M cyclic AMP (Sigma) and 80 mM NaCl. The appropriate CAP:DNA ratio for each mutant was determined by titration. After a 10-min incubation ( $\sim 23^\circ\text{C}$ ), the samples were loaded on to a native 10% polyacrylamide gel (75:1 w/w acrylamide:bis) and electrophoresed in 1/2 TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.3) and 20- $\mu$ M cAMP at 420 V in a constant-temperature gel box (Biorad) at  $25^\circ\text{C}$ .



**Fig. 3** Experimentally measured mobilities for the site A+D mutant complexes plotted against mobilities calculated from the linear least squares analysis. The quality of the fit is represented in terms of  $r^2$ , the fraction of the total sum of squares that is accounted for by the linear regression of  $\mu_{\text{rel}}$  on independent variables corresponding to the sequences and positions of binding site sequence elements. Error associated with measurement of  $\mu_{\text{rel}}$  estimated as  $\pm 0.015$ . *a*, Dinucleotide model,  $r^2 = 0.965$  (corrected for degrees of freedom), r.m.s. =  $\pm 0.018$ . *b*, Mononucleotide model  $r^2 = 0.841$  (corrected for degrees of freedom), r.m.s. =  $\pm 0.040$ . See M.R.G., J. Nadeau and D.M.C. (manuscript in preparation) for details.

visual inspection of the large data set, it is clear that a correlation exists between binding site sequence and protein-induced bending (see Table 1 for representative examples). Maximal bending occurs when a site centred  $\sim 10.5$  bp from the binding site dyad is A-T rich. Minimal bending occurs when this site is G-C rich. The reverse pattern, though less pronounced, is observed at a site centred roughly 16 bp from the dyad axis. Weaker correlation exists for the nucleotides joining the two sites. We infer that the sites 10.5 and 16 bp from the binding site dyad function as bending loci. A quantitative analysis presented below supports this conclusion.

CAP shares the helix-turn-helix recognition motif of many other regulatory DNA binding proteins<sup>33-35</sup>. Correct alignment of the recognition helices requires that the minor groove face the protein at the dyad axis of the binding site<sup>7</sup>. With a helical repeat of  $\sim 10.5$  bp the minor groove faces the protein at the first bending locus 10.5 bp away and the major groove faces the protein at the second bending locus 16 bp away. Bending primarily by roll into the minor groove at the first site and the major groove at the second site explains the maximal dependence of mobility on the nucleotide sequence of these loci. We conclude that A-T-rich sequences favour minor groove compression, whereas G-C-rich sequences favour major groove compression in the CAP-DNA complex.

## Dinucleotide ranking

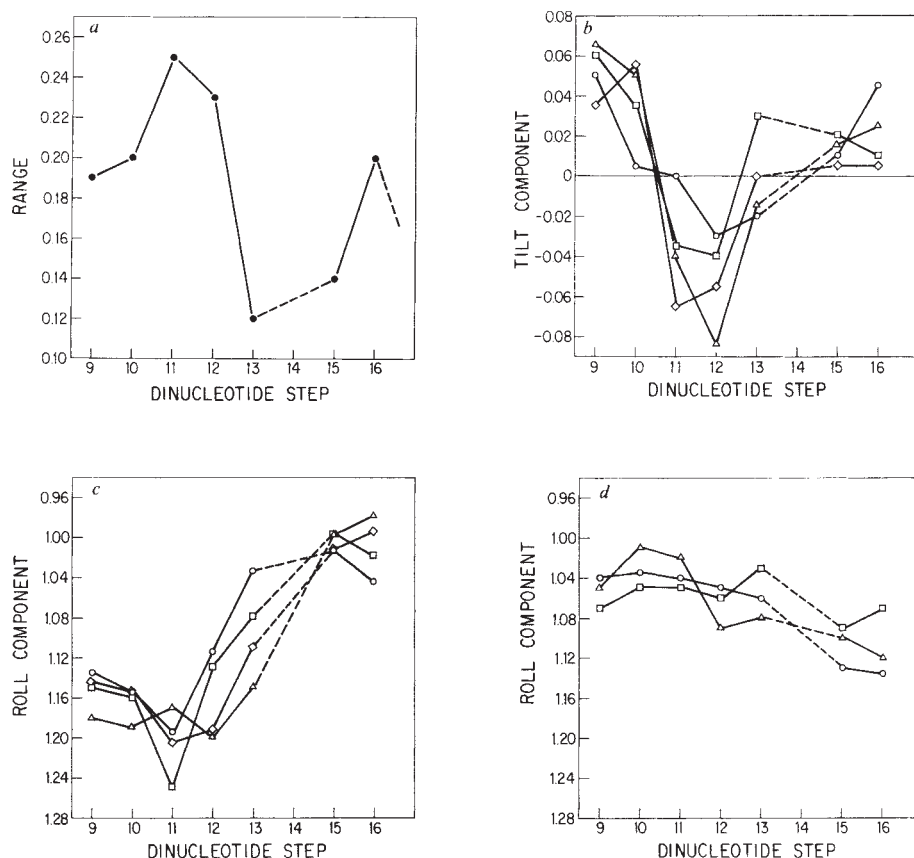
Though clear sequence patterns emerge from visual inspection of the data (see Table 1), a ranking of sequence elements according to their influence on bending requires further quantitative treatment (M.R.G., J. Nadeau and D.M.C., manuscript in preparation). The dependence of an observable DNA property (in this case the gel mobility of the CAP complex) on sequence can be expressed by a set of linear equations, one for

each mutant examined, with independent variables or 'parameters', whose values are related to the influence contributed by particular sequence elements at given positions in the molecule<sup>36</sup>. The accuracy of the representation depends on the identification of appropriate sequence elements that contribute linearly to the observable property. Linear least squares analysis is used to optimize the fit of the data to various models. In all cases the number of mutants studied is substantially greater than the number of independent variables, so the derived parameters do not necessarily predict the observed properties of any one mutant precisely.

Figure 3*a* plots experimentally measured complex mobilities versus those predicted by linear least squares analysis of the dinucleotide model. The scatter of the data around the predicted lines is comparable to the estimated error in measuring the data points. The success of this model confirms the inherent assumption that the dinucleotide constitutes a structurally independent bending unit to the degree of accuracy permitted by our data set. For contrast, in Fig. 3*b* we show that regression based on mononucleotides yields less satisfactory results, with scatter of the observations around the predicted line nearly 3 times larger than the estimated standard error in the data. In view of the success of the dinucleotide model and the limited size of the data set, we did not investigate a trinucleotide model.

Models based on overlapping elements (dinucleotides) introduce an interdependency of successive sequence elements which reduces the number of independent variables below the number of dinucleotides, and therefore prevents direct ranking of dinucleotides from parameter values (see ref. 37 and M.R.G., J. Nadeau & D.M.C., manuscript in preparation). Instead we have used the least-squares-determined parameters to predict the mobilities of all CAP-DNA complexes, and averaged the calculated mobilities over bases W and Z in the tetranucleotide sequences WpXpYpZ. This yields an estimate of the bending

**Fig. 4** *a*, Range of dinucleotide parameter estimates,  $\bar{\mu}_{\max} - \bar{\mu}_{\min}$ , as a function of binding site position. Insufficient data prevented the ranking of dinucleotides at positions 14 and 17; nevertheless from the mobilities that could be predicted it is clear that neither position represents a local maximum. *b*, Tilt component of Pu-Pu·Py-Py dinucleotide wedges as a function of binding site position. Tilt is defined as one half of the difference of complementary dinucleotide parameter estimates.  $\circ$ ,  $(\bar{\mu}_{AA} - \bar{\mu}_{TT})/2$ ;  $\Delta$ ,  $(\bar{\mu}_{GA} - \bar{\mu}_{TC})/2$ ;  $\square$ ,  $(\bar{\mu}_{AG} - \bar{\mu}_{CT})/2$ ;  $\diamond$ ,  $(\bar{\mu}_{GG} - \bar{\mu}_{CC})/2$ . *c*, Roll component of G-C rich dinucleotide wedges as a function of binding site position. Roll is defined as the average of complementary dinucleotide parameter estimates.  $\circ$ ,  $(\bar{\mu}_{CA} + \bar{\mu}_{TG})/2$ ;  $\Delta$ ,  $\bar{\mu}_{GC}$ ;  $\square$ ,  $\bar{\mu}_{CG}$ ;  $\diamond$ ,  $(\bar{\mu}_{GG} + \bar{\mu}_{CC})/2$ . Roll components reach local maxima (low  $\bar{\mu}$ -values) or minima (high  $\bar{\mu}$ -values) at bend centres depending on the preferred direction of bending. *d*, Roll component of A-T-rich dinucleotide wedges as a function of binding site position.  $\circ$ ,  $(\bar{\mu}_{AA} + \bar{\mu}_{TT})/2$ ;  $\Delta$ ,  $\bar{\mu}_{AT}$ ;  $\square$ ,  $\bar{\mu}_{TA}$ .



influence of the internal dinucleotides XpY at a fixed position (see Table 2 legend). The averaging procedure dampens context effects and results in parameters,  $\bar{\mu}$ , that approximate the contribution of each dinucleotide wedge to the overall bend.

The  $\bar{\mu}$ -parameters confirm the qualitative observations stated previously (see Table 2): the range of the parameters varies from one position to the next in a sinusoidal fashion. Maximal amplitude is reached at positions 11 and 16 bp from the binding site dyad axis (see Fig. 4*a*), where the minor and major grooves face the protein in succession. This result shows quantitatively that the sequence dependence of bending is maximal at those sites. The ranking orders of dinucleotides in positions 10 and 11 are roughly opposite those in positions 15 and 16 and coincide with the observation that A-T-rich sequences favour bending where the minor groove faces the protein and G-C-rich sequences favour bending where the major groove faces the protein. The similarity of these observations to the DNA bendability rules deduced for nucleosomes<sup>19,20</sup> strongly implies that this phenomenon is general and will apply to other complexes where non-specific protein-DNA interactions lead to DNA bending.

## Roll and tilt components

Certain sequence dependences emerge from our analysis that were not observed in studies of nucleosomal DNA<sup>19,20,27</sup>. In particular, the  $\bar{\mu}$ -parameters for complementary dinucleotides are not equivalent, particularly for Pu-Pu·Py-Py pairs, where Pu represents a purine and Py, a pyrimidine. This can be explained if both roll and tilt bending contribute to form dinucleotide wedges near the bend centres<sup>38</sup>: the difference in  $\bar{\mu}$ -parameters for complementary dinucleotides reflects an anisotropic bendability about the tilt axis. Formation of a wedge in the overall bend direction requires a tilt component unless the roll axis is perpendicular to the bending plane. (The earliest analyses of chromatin nucleotide sequence periodicities were interpreted as indicative of tilt bending<sup>38</sup>.) To examine this feature, we deconvoluted the wedge values into roll and tilt

**Table 2** Dinucleotide rankings based on  $\bar{\mu}$ -parameters

| Bend towards minor groove<br><i>n</i> = 10 |           | Bend towards major groove<br><i>n</i> = 16 |
|--------------------------------------------|-----------|--------------------------------------------|
| <i>n</i> = 11                              |           |                                            |
| best benders                               |           |                                            |
| AT = 1.01                                  | AT = 1.02 | GC = 0.98                                  |
| TT = 1.03                                  | AA = 1.04 | CC = 0.99                                  |
| AA = 1.04                                  | TT = 1.04 | GG = 1.00                                  |
| TC = 1.04                                  | TA = 1.05 | CG = 1.02                                  |
| TA = 1.05                                  | AG = 1.09 | GT = 1.02                                  |
| CT = 1.07                                  | AC = 1.10 | TC = 1.03                                  |
| CC = 1.10                                  | GT = 1.11 | TG = 1.03                                  |
| GT = 1.12                                  | GA = 1.12 | CT = 1.04                                  |
| TG = 1.13                                  | GG = 1.14 | CA = 1.06                                  |
| GA = 1.14                                  | CT = 1.16 | AG = 1.06                                  |
| AG = 1.14                                  | CA = 1.16 | AC = 1.06                                  |
| AC = 1.14                                  | GC = 1.17 | TA = 1.07                                  |
| CG = 1.16                                  | TC = 1.20 | GA = 1.08                                  |
| CA = 1.18                                  | TG = 1.23 | TT = 1.09                                  |
| GC = 1.19                                  | CG = 1.25 | AT = 1.12                                  |
| CG = 1.21                                  | CC = 1.27 | AA = 1.18                                  |
| worst benders                              |           |                                            |

Each dinucleotide parameter,  $\bar{\mu}$ , corresponds to the average mobility of sequences containing the dinucleotide wedge:  $\bar{\mu}(n) = \sum_w \sum_z WXYZ/m$ , where *n* is the position to be evaluated and *m* is the number of calculated mobilities contributing to the average. See M.R.G., J. Nadeau & D.M.C. (manuscript in preparation) for details. In testing the use of these parameters, we found, for example, that addition of the four position-dependent  $\bar{\mu}$ -values for a pentanucleotide centred at the 10-11 bp spaced bending locus gave excellent correlation (r.m.s. of a plot analogous to Fig. 3*a* = ±0.025) with the mobilities calculated from the full overlapping dinucleotide model.



components, defined respectively as the average and one half the difference of the  $\bar{\mu}$ -parameters of complementary dinucleotides; these values are plotted as a function of binding site position in Fig. 4b–d. The tilt component for all Pu–Pu·Py–Py dinucleotide pairs changes sign between positions 10 and 11 as expected when passing through a position where the roll axis is perpendicular to the plane of bending. From the sign of the tilt component and the requirement that the minor groove face the protein at the dyad axis, we deduce that the wedge narrows toward the Py–Py element of Pu–Pu·Py–Py dinucleotides. This leads to a preference for Pu–Pu dinucleotides over Py–Py on the 5' side of the centre of a bend toward the minor groove, with the reverse preference on the 3' side.

Systematic change of the sign of the Pu–Pu·Py–Py tilt components around the minor groove roll point supports the view, also justified by statistical significance tests (M.R.G., J. Nadeau and D.M.C., manuscript in preparation), that models which consider the complementary dinucleotides to be equivalent account significantly less well for the data. This seems not to be the case with the nucleosome data set<sup>19,20</sup>, for unknown reasons.

Beyond position 12, the tilt function no longer behaves in a sufficiently systematic way to enable us to deduce tilt components in a bend toward the major groove. It should be noted, however, that tilt need not be in the same direction as at a minor groove bend because of possible coupling between roll and tilt.

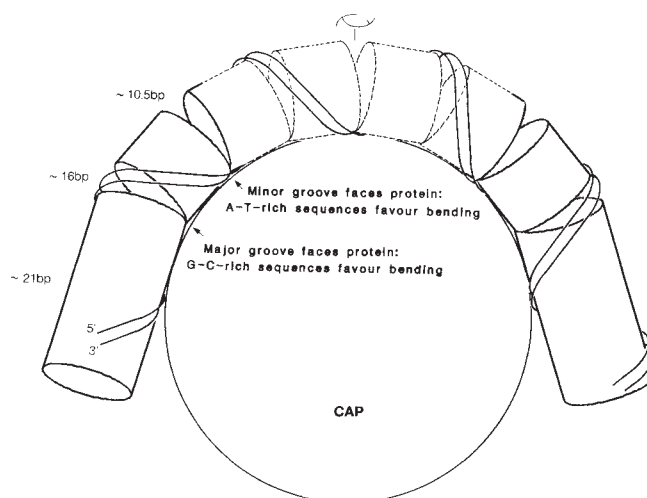
Roll components reach local maxima (low  $\bar{\mu}$ -values) or minima (high  $\bar{\mu}$ -values) at bend centres depending on the preferred direction of bending. For G–C-rich wedges a maximum occurs near position 16, where the major groove faces the protein, and for A–T-rich wedges a maximum occurs near position 11, where the minor groove faces the protein.

### Extent of the bending domain

The emergence of bending loci spaced every 5–6 bp led us to investigate a putative third bending site spaced 21–22 bp from the dyad. All of the mutations generated at this position (site E) fail to alter the mobility of the CAP–DNA complex significantly, with the exception of those containing an A<sub>5</sub> intrinsic bending locus centred there (similar to mutant A22, Table 1). This result puts an upper limit on the size of the bending domain: if we assume twofold structural symmetry, it encompasses less than 42 bp of DNA. It seems likely that the same limit also applies to the region of protein–DNA contact. Changes in sequence 17 bp from the dyad affect the mobility of the complex. In a structural sense, this determines the lower limit of the binding domain: the site spans at least 34 bp.

### Binding strength versus bending angle

The DNA sequence dependence of protein–DNA complex mobility indicates that some binding sites bend more easily than others. The free energy required for DNA bending reduces the favourable free energy of the binding interaction. Furthermore, a reduction in DNA bending results in the weakening of favourable protein–DNA contacts which stabilize the complex. To test for a correlation between binding strength and induced bending angle, we measured the relative binding affinities of selected bending mutants by the gel electrophoretic method<sup>39,40</sup>. DNA fragments containing mutant binding sites were independently mixed with a shorter wt binding site fragment and titrated with CAP. The partition-equilibrium distribution of products was resolved by gel electrophoresis. The ratios of bound and free DNAs were determined, and the ratio  $K_{rel}$  of mutant binding constant  $K_n$  to  $K_{wt}$  for the wild type was calculated from the data. The results collected in Table 1 indicate that mutations at bending site 1 which decrease the extent of bending also reduce the binding affinity, in some cases by more than an order of magnitude. This correlation persists at site 2, though it is less pronounced. These results show that the relative preference for DNA bending by different sequences is accompanied by a



**Fig. 5** Schematic representation of the CAP–DNA complex. Bending results from the phased groove compression of non-consensus DNA interacting electrostatically with the surface of CAP. Here, this is represented by a hinged rod wrapped around a sphere. A double helix is superimposed on the rod to illustrate the positions of the grooves relative to the protein surface. In non-specifically interacting regions, bending loci are spaced by 5–6 bp and coincide with positions where the minor and major grooves face the protein. Regions of DNA interacting electrostatically with a charged protein surface join the bending loci. A–T-rich sequences favour bending and binding where the minor groove faces the protein (~10.5 bp from the pseudo-dyad axis) and G–C-rich sequences favour bending and binding where the major groove faces the protein (roughly 16 bp from the pseudo-dyad axis). The bending site spans at least 34 bp, but less than 42 bp. The DNA sequence of the consensus domain also modulates DNA bending but is subject to the additional constraints of sequence-specific recognition (M.R.G. and D.M.C., unpublished results). As such, the rules elucidated from mutagenesis of the non-consensus domain do not apply. For visual clarity, however, the hinged rod extends through the consensus region in this representation (dotted lines).

significant and correlated variation in the binding affinity of a regulatory protein–DNA complex. This conclusion, along with the finding of similar bending and binding by a variety of sequences, strengthens the hypothesis that protein–DNA interactions in the distal binding domain are non-specific, probably electrostatic<sup>8</sup>.

It is likely that the weaker correlation of binding strength and bending angle at the second site originates from the relative positions of the two bending loci. The internal site (10.5 bp) affects the helix trajectory of a large segment of distal DNA and thus the stabilizing electrostatic interactions it can make. The external site (16 bp), situated beyond the strong contacts specified by the internal site, also influences helix trajectory, but its proximity to the binding site boundary restricts the number of significant distal interactions it can promote.

### Range of bending angles

To calibrate absolute changes in the induced bending angle caused by mutation, we compared the mobilities of selected mutants to those of wt fragments whose end-to-end distance had been modulated by the incorporation of intrinsic bends of known magnitude (see Fig. 2). For standards, we fused single CA<sub>6</sub>C tracts, in-phase and out-of-phase, with the wt CAP-induced bend, each tract deflecting the helix trajectory by ~20 degrees<sup>41,42,47</sup>. Optimum phasing was found for an A tract centred at 22 bp from the dyad (mutant A22), as expected for effective bending of the A tract toward the minor groove at its centre<sup>32</sup>. A plot of the bending angle deviation versus relative mobility for the three standards, +20 degrees (A22), 0 degrees (wt), –20 degrees (A27, with an out-of-phase A-tract bend), displays a nearly linear relationship in which mobility increases

as bending decreases in the range of bending angles examined (data given in Table 1). The relative mobilities of mutants were used to extrapolate the changes in bending angles from this curve (see Table 1). We conclude that sequence changes in the CAP binding site open the bending angle by as much as 1.6 A-tract equivalents (32 degrees, mutant symr) and close it by no more than 0.6 A-tract equivalents (13 degrees, mutant sym1).

## Conclusion

It now seems likely that phasing of groove compressing sequences in non-specifically interacting regions (as shown in Fig. 5) will prove to be a general feature of DNA bent by protein binding, as similar results are found for such functionally and biologically divergent systems as CAP protein and nucleosomes. It should be noted, however, that modulation of protein-induced bending can also result from geometric constraints imposed by hydrogen-bonding and other sequence-specific contacts<sup>14,16</sup>, close association of protein binding sites with A-tract bending sites<sup>10-12</sup>, and the influence of other DNA binding proteins and drugs<sup>43,44</sup>. The consensus sequence of the CAP binding site also influences the extent of bending (ref. 1 and M.R.G. and D.M.C., unpublished results).

Received 30 November 1987; accepted 20 April 1988.

- Fried, M. G. & Crothers, D. M. *Nucleic Acids Res.* **11**, 141-158 (1983).
- Kolb, A., Spassky, A., Chapon, C., Blazy, B. & Buc, H. *Nucleic Acids Res.* **11**, 7833-7852 (1983).
- Wu, H.-M. & Crothers, D. M. *Nature* **308**, 509-513 (1984).
- Kotlarz, D., Fritsch, A. & Buc, H. *EMBO J.* **5**, 799-803 (1986).
- Dripps, D. & Wartell, R. M. *J. Biomolec. struct. Dyn.* **5**, 1-13 (1987).
- Porschke, D., Hillen, W. & Takahashi, M. *EMBO J.* **3**, 2873-2878 (1984).
- Weber, I. T. & Steitz, T. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3973-3977 (1984).
- Waricker, J., Engleman, B. P. & Steitz, T. A. *Proteins* **2**, 283-289 (1987).
- Shuey, D. J. & Parker, C. S. **323**, 459-462 (1986).
- Kunke, G., Fritz, H.-J. & Ehrling, R. *EMBO J.* **6**, 507-513 (1987).
- Mukherjee, S., Parel, I. & Bastia, D. *Cell* **43**, 189-197 (1985).
- Ryder, K., Silver, S., Delucia, A. L., Fanning, E. & Tegtmeyer, P. *Cell* **44**, 719-725 (1986).
- Zahn, K. & Blattner, F. R. *Science* **236**, 416-422 (1987).
- Hatfull, G. F., Noble, S. Z. & Grindley, N. D. F. *Cell* **49**, 103-110 (1987).
- Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. *Nature* **311**, 532-537 (1984).
- McClarín, T. A. *et al. Science* **234**, 1526-1541 (1986).
- Anderson, J. E., Ptashne, M. & Harrison, S. C. *Nature* **326**, 846-852 (1987).
- Liu-Johnson, H.-N., Gartenberg, M. R. & Crothers, D. M. *Cell* **47**, 995-1005 (1986).
- Drew, H. R. & Travers, A. A. *J. molec. Biol.* **186**, 773-790 (1985).
- Satchwell, S. C., Drew, H. R. & Travers, A. A. *J. molec. Biol.* **191**, 659-675 (1986).
- Koudelka, G. B., Harrison, S. B. & Ptashne, M. *Nature* **326**, 886-888 (1987).
- Hogan, M. E. & Austin, R. H. *Nature* **329**, 263-266 (1987).
- deCrombrughe, B., Busby, S. & Buc, H. *Science* **224**, 831-838 (1984).
- Majors, J. thesis, Harvard Univ. (1977).
- Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. *Nature* **311**, 232-235 (1984).
- Ebright, R. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 6083-6087 (1987).
- Calladine, C. R. & Drew, H. R. *J. molec. Biol.* **192**, 907-918 (1986).
- Travers, A. & Klug, A. *Nature* **327**, 280-281 (1987).
- Wells, J. A., Vasser, M. & Powers, D. B. *Gene* **34**, 315-323 (1985).
- Lerman, L. S. & Frisch, H. L. *Biopolymers* **21**, 995-997 (1982).
- Lumpkin, O. J. & Zimm, B. H. *Biopolymers* **21**, 2315-2316 (1982).
- Zinkel, S. S. & Crothers, D. M. *Nature* **328**, 178-181 (1987).
- McKay, D. B. & Steitz, T. A. *Nature* **290**, 744-749 (1981).
- Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *Nature* **290**, 754-758 (1981).
- Pabo, C. O. & Lewis, M. *Nature* **298**, 443-447 (1982).
- Stormo, G. D., Schneider, T. D. & Gold, L. *Nucleic Acids Res.* **14**, 6661-6679 (1986).
- Gray, D. M. & Tinoco, I. Jr *Biopolymers* **9**, 223-244 (1970).
- Trifonov, E. N. *CRC Crit. Rev. Biochem.* **9**, 89-106 (1985).
- Fried, M. G. & Crothers, D. M. *Nucleic Acids Res.* **9**, 6505-6525 (1981).
- Garner, M. M. & Revzin, A. *Nucleic Acids Res.* **9**, 3047-3060 (1981).
- Levene, S. D., Wu, H.-M. & Crothers, D. M. *Biochemistry* **25**, 3988-3995 (1986).
- Koo, H.-S. thesis Yale Univ. (1987).
- Drew, H. R. & Calladine, C. R. *J. molec. Biol.* **195**, 143-173 (1987).
- Low, L., Drew, H. R. & Waring, M. *Nucleic Acids Res.* **14**, 6785-6801 (1986).
- Malan, T. P., Kolb, A., Buc, H. & McClure, W. R. *J. molec. Biol.* **180**, 881-909 (1984).
- Buc, H. *et al. in RNA Polymerases and the Regulation of Transcription* (eds Reznikoff, W. S. *et al.*) 115-125 (Elsevier, Amsterdam, 1987).
- Griffith, J., Bleyman, M., Rauch, C. A., Kitchin, P. A. & England, P. A. *Cell* **46**, 717-724 (1986).

## LETTERS TO NATURE

### Do massive black holes reside in elliptical galaxies?

A. C. Fabian\* & C. R. Canizares†

\* Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK  
† Center for Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Evidence that radio-loud quasars and powerful radio galaxies occur in elliptical galaxies<sup>1,2</sup> suggests that most bright ellipticals now contain a 'dead' or inactive quasar<sup>3,4</sup>. In most theories, the central 'engine' is a massive, perhaps rapidly spinning black hole<sup>5</sup>, with a mass  $\sim 10^8$ - $10^9 M_\odot$ , deduced from simple energetic arguments or from the properties of the broad-line emitting regions<sup>6</sup>. There have been a number of searches for the massive black holes that are expected to reside at the centres of elliptical galaxies<sup>7-9</sup>. Here we examine the accretion by a central black hole of the hot interstellar medium in an elliptical galaxy<sup>10,11</sup>, and estimate the minimum expected luminosity and the manner of its emission. It is not obviously detected at any wavelength. Contrary to the problem raised by Gunn<sup>12</sup> of 'feeding the monster', we raise the problem of 'starving' it, if indeed there is a monster.

The interstellar medium of elliptical galaxies was discovered<sup>10,13,14</sup> using the Einstein Observatory. Stellar mass-loss is trapped in the potential well of the galaxy and cools there into new stars. Unlike gas in a spiral galaxy, which is rotationally supported, the gas in an elliptical is supported by its thermal pressure. As the gas temperature is  $\sim 10^7$  K, the pressure exceeds  $10^6 \text{ cm}^{-3} \text{ K}$  within the innermost 100 pc (ref. 15). Accretion of the gas on to a central black hole has been discussed<sup>10,11</sup>. The role of cooling in the gas has been considered and what is essentially an upper-limit to the steady accretion luminosity has been deduced by setting the cooling time equal to the free-fall time at the accretion radius of  $L_a \approx 2.10^{45} M_8 T^{3/2} \text{ erg s}^{-1}$ . The mass of the hole is  $M = 10^8 M_8 M_\odot$ , the gas temperature at the accretion radius is  $10^7$  K, the cooling function is approximated by  $10^{-19} T^{-1/2} \text{ erg cm}^3 \text{ s}^{-1}$ , and a 10% matter-to-energy conversion efficiency is assumed. Such a high luminosity is clearly not present in most nearby elliptical galaxies.

A lower limit is obtained by using the classic Bondi<sup>16</sup> formula for accretion by a point mass in a static medium. The situation of a black hole in an elliptical galaxy is likely to be a close approximation to this and is perhaps the only reasonable example. Taking the accretion radius as  $R_a = \alpha GM/c_s^2$ , where the sound speed in the gas,  $c_s \approx 10^4 T^{1/2} \text{ cm s}^{-1}$  and  $\alpha$  is a factor including the ratio of specific heats and effects of cooling ( $\alpha > 0.5$



and is probably  $\sim 1$ ), we obtain  $L_b = 4.10^{41} \alpha^2 P_6 T_7^{-5/2} M_8^2 \text{ erg s}^{-1}$ ,  $P$  being the thermal pressure of the gas ( $P = 10^6 P_6 \text{ cm}^{-3} \text{ K}$ ).  $L_b$  is much less than  $L_a$  in the parameter range of interest, but is still high enough to be easily detectable if radiated in the radio, infrared, optical, near-ultraviolet or X-ray wavebands. For example, the total X-ray luminosity of the gas in elliptical galaxies<sup>17</sup> (which is well resolved spatially in nearby cases) lies between  $10^{40}$  and a few times  $10^{41} \text{ erg s}^{-1}$ .  $M_8 = 1$  is probably a lower limit; the gas temperature is unlikely to exceed  $10^7 \text{ K}$  by much, and could well be considerably less if cooling is rife in the core of the galaxy (then  $L_b \rightarrow L_a$ ).

We are thus led to the conclusion that a black hole of mass  $10^8$ – $10^9 M_\odot$  in the centre of most elliptical galaxies would be readily detected by the accretion luminosity if the radiative efficiency in X-rays of the inflow exceeds about 0.1–1%. There is no problem in explaining the powering of M87 (the X-ray luminosity of the central source<sup>18</sup> is  $10^{43} \text{ erg s}^{-1}$ , which, if equated to  $L_b$  with  $P_6 = T_7 = 1$  (ref. 19), implies a black hole mass of  $10^9 M_\odot$ ) or indeed of any bright active elliptical. What is of concern here is the low central luminosity of most normal ellipticals. We exclude the ellipticals in rich clusters which, except for slow-moving central galaxies, are kept devoid of gas by ram-pressure stripping.

We have derived upper limits to the mass of any central black hole from X-ray observations of the six galaxies<sup>20</sup> (Table 1). An

**Table 1** Upper limits to mass of any central black hole from X-ray observation of the 6 galaxies

| Galaxy<br>NGC | $M_B$ | Black hole<br>$10^8 M_\odot$ | Mass<br>$10^8 M_\odot$ |
|---------------|-------|------------------------------|------------------------|
| 720           | -21.5 | 1.3                          | 6.0                    |
| 1395          | -21.2 | 1.2                          | 6.6                    |
| 4382          | -22.0 | 1.2                          | 6.2                    |
| 4472          | -22.8 | 0.26                         | 0.48                   |
| 4636          | -21.6 | 0.50                         | 1.0                    |
| 4649          | -22.2 | 0.44                         | 0.86                   |

estimate for the central gas density (and thus pressure) can be related to its luminosity assuming the standard 'King-law' profile for the surface brightness profile (see equation 6 of ref. 17). This leads to

$$\frac{L_b}{L_X} \approx \frac{M_8^2}{2} \left( \left( \frac{L_X}{10^{41} \text{ erg s}^{-1}} \right) a^3 T_7^3 \right)^{-1/2}$$

where  $L_X$  is the total X-ray luminosity of the galaxy due to gas,  $a$  is the core radius of the gas distribution in kpc and  $\alpha = 0.5$ . On the assumption that the accretion luminosity appears as X-rays, we derive an upper limit on the black hole mass by setting  $L_b$  equal to the maximum possible luminosity of a point source. This is estimated from the surface brightness profiles of ref. 20, accounting for the spatial resolution of the detector (for NGC4382, all the central power could be from a point source; the other five show no evidence for a point component, so we arbitrarily use half the central power as an upper limit). The third column in Table 1 gives the limits of  $M_8$  using the best estimates of  $a$  and  $T$  from the fits<sup>20</sup> (the first three galaxies were observed with the IPC detector which was of lower resolution and could not directly measure  $a$ , so we assumed  $a = 1$ , close to the values measured with the High Resolution Imager for the other three galaxies). This is nevertheless an upper limit on  $M$  because of our conservative assumptions about the X-ray luminosity, the neglect of any possible pressure rise within the inner kpc of the galaxy (see the trends in ref. 13) and the neglect of cooling. The fourth column gives a still more conservative upper limit derived using 90% confidence upper limits to  $a$  and  $T$  (we had to assume a limit of 5 on  $a$  for NGC4382 and a limit of 3 on  $T$  for NGC4382 and NGC720). In this case, the galaxies observed with the IPC give considerably higher limits to the

black hole mass. The mass limits scale with Hubble constant as  $H_0^{-5/4}$  (we use  $50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ).

Given our assumptions, we can rule out the presence of a black hole mass of a few times  $10^7 M_\odot$  in NGC4472, which is the brightest elliptical in the Virgo Cluster, and our reasonable limits for the other five galaxies are still below  $1.5 \times 10^8 M_\odot$ . Note (Table 1) that all of these galaxies are brighter than  $M_B = -21.2$ , and three are brighter than  $-22.0$ .

Our limits can be compared with estimates based on the arguments first set forth by Soltan<sup>3</sup>. Using conservative values for quasar number counts, Soltan finds  $\sim 4.7 \times 10^4 \eta_{0.1} M_\odot \text{ Mpc}^{-3}$  for the mass accumulated in black holes, where the mass-to-energy conversion efficiency is  $0.1 \eta_{0.1}$  (Rees<sup>21</sup> quotes a value of  $3.10^4$ ). The typical mass expected in a single remnant black hole depends on how faint a host galaxy one is willing to allow (note that galaxies at least as faint as  $M_B = -22.0$  must be involved simply to account for the estimated density of quasars at  $z = 2.5$ ; ref. 22). For example, we find  $2.0 \times 10^9 \eta_{0.1} h_{50}^3 M_\odot$  if quasars reside in all galaxies brighter than  $M_B = -22.0$  and  $1.8 \times 10^8 \eta_{0.1} h_{50}^3 M_\odot$  if the cutoff is  $-21.2$  (we use the galaxy luminosity function of J. Huchra and R. Burg, personal communication; and  $H_0 = 50 h_{50} \text{ km s}^{-1} \text{ Mpc}^{-1}$ ). The numbers change little if it is assumed that ellipticals house radio-loud quasars and spiral galaxies the radio-quiet ones as the elliptical to spiral ratio is similar to the radio-loud to radio-quiet one. The independent analysis of Wandel and Mushotzky<sup>6</sup> also points to remnant quasar masses  $> 10^8 M_\odot$ , as do some of the arguments in Rees<sup>21</sup> (Seyfert nuclei can have lower masses but these did not enter in Soltan's estimate), and recent work by Smith *et al.*<sup>4</sup> finds that host galaxies are indeed predominantly brighter than  $-21.5$ .

Thus there is a discrepancy between the upper limits for the black hole masses we infer for nearby bright ellipticals, and the masses estimated from considerations of the density and properties of the quasars themselves, which should be considered as lower limits because the assumptions were conservative (for  $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$  the lower limits decrease by 2.4; this narrows but does not eliminate the discrepancy). The two estimates could be reconciled if one accepts that much fainter galaxies also hosted quasars, so the average remnant mass is less than  $10^7 M_\odot$ , although one must then also reject the findings of Wandel and Mushotzky<sup>6</sup> and Smith *et al.*<sup>4</sup>. If instead one attempts reconciliation by arguing that only a fraction of bright galaxies host quasars, then these will have remnant masses  $> 10^9 M_\odot$ , values of  $L_b > 10^{43} \text{ erg s}^{-1}$  and raise problems concerning the observed number density of quasars. Presumably these galaxies still show nuclear activity and are classified as 'active' in some way.

Before accepting these conclusions or rejecting the black hole hypothesis for quasars, which seems to be the only generally accepted idea<sup>21</sup> that can account for the gross energy budget (but not the details) of quasars and of the rapid variability that many objects display, we examine some of the loopholes in the above arguments.

First, it is possible that the accretion luminosity is being produced but that it is in some inaccessible part of the spectrum. For instance, blackbody emission from a massive black hole emerges predominantly in the extreme ultraviolet which is unobservable because of photoelectric absorption by galactic hydrogen. Such emission is seen by soft X-ray instruments in some Seyfert galaxies<sup>23,24</sup> and has been hypothesized to exist in some ellipticals<sup>25</sup>. X-rays from accretion on to a black hole could be absorbed by the surrounding gas and so not be observable. The energy must still be emitted however, probably in the infrared band. This could account for some of the Infrared Astronomical Satellite luminosity of elliptical galaxies (see, for example, ref. 26). The power may alternatively be in kinetic energy as jets. If the jet power is not dissipated on traversing the galaxy and has no 'working surface', such as intracluster gas, then there may be little electromagnetic radiation. (For example, what would NGC1265 have looked like before it fell into the Perseus

Cluster?) In support of this possibility, we note that many ellipticals do have some radio source (often of apparently low power) in their cores. This leads to the interesting consequence that intergalactic space is being filled with relativistic plasma. It is also possible that the radiation is highly beamed, so that most ellipticals are BL Lac objects seen off-axis. Otherwise the power may be radiated as neutrinos or some other particles, but there is no current accretion theory that suggests it should be.

Second, the interstellar medium may not penetrate to the hole. Back-pressure from very hot, outflowing or relativistic gas around the hole may be important. This must not radiate at luminosities above  $\sim 0.1 L_b$ , however, which seems unlikely. A very massive star cluster<sup>27</sup>, with a density exceeding  $3.10^7 M_\odot \text{pc}^{-3}$ , surrounding the hole could raise the gas temperature to above a few times  $10^7 \text{K}$  and so reduce the Bondi-rate on to the hole. Such a cluster would, however, have a short relaxation time and so be problematic. Disruption of stars in the neighbourhood of a massive black hole could also lead to gas which accretes and so enhances the luminosity. This has been discussed by Ozernoy *et al.*<sup>29</sup>, and the effect has been used to limit the masses of black holes in galactic nuclei. A major difficulty<sup>30</sup>, then, is the estimation of the rate at which the stellar orbits are repopulated once those stars that can fall close to the hole have been disrupted and their gas captured. Back pressure from a high velocity wind could prevent accretion of the hot medium by a black hole. Our galactic nucleus has such a wind, which has itself been used to limit the possible mass of a central black hole<sup>31</sup>.

The gas could also be prevented from reaching the hole by rotation. If the diffuse gas, which is also cooling, has sufficient angular momentum to orbit the hole it may form a disk or torus. Observations of mass transfer in binary stars suggests that the angular momentum is then transported outward by viscosity and most of the gas flows inward and so will be accreted. The accretion luminosity will be substantially below  $L_b$  if most of the hot gas at  $R_a$  is centrifugally supported, or if viscosity is much lower than in stellar-mass accretion disks. Some accretion must still persist along the rotation axis. There is evidence for cold, rotating gas disks in some elliptical galaxies (see ref. 32 for example) but these do not generally show the same rotation axis as the stars. Most of the diffuse gas in the core of an elliptical will be from stellar mass-loss and so have a similar angular momentum to the stars. The mass accretion rate  $L_b$  corresponds to  $\sim 10^{-5} M_\odot \text{yr}^{-1}$  for  $M_8 \sim 1$  and is supplied by the innermost  $10^6 M_\odot$  of stars. These stars will lie close to, or within,  $R_a$  and so if their gas is not accreted it must either be blown out of the central region (note that the presence of the hot interstellar medium in ellipticals means that the central gas cannot be blown out of the galaxy) or have sufficient angular momentum to orbit the hole and the viscosity must be much reduced. We have argued elsewhere<sup>10</sup> that angular momentum transport will occur in the hot gas. We doubt, however, that the flow near the hole is highly radial, when a low efficiency is possible<sup>33</sup>. Soltan's limits are more sensitive to  $\eta$  than ours, so lowering the efficiency used in our calculations makes the inconsistency worse. To resolve the discrepancy, one must argue that the efficiency is lower now, but was higher at earlier epochs (then the demise of quasars is partially due to a change in efficiency). Note that any situation where most of the accretion flow penetrates to about 10 Schwarzschild radii of the hole (or indeed any other kind of massive compact object) in a non-radial manner should release most of the gravitational energy.

Finally, it is possible that the flow is time-dependent and an accretion episode produces so much back-pressure from heating or radiation that it stems further inflow for some period. As the interstellar gas is already highly ionized, it is not clear that heating will be successful in this way. There is no sign of such events, which should recur on the free-fall time from the accretion radius of  $\sim 10^3 \text{yr}$ . The fraction of time that the nucleus

is active must also be very short (and presumably very powerful) in order that most galaxies are seen to be inactive.

In summary, we find that one simple test for massive black holes in the centres of bright elliptical galaxies fails by predicting accretion luminosities about one order of magnitudes higher than observed. Another test for black holes has been proposed by Gerhard<sup>34</sup> and involves the destruction of triaxiality by scattering of stars in the central cusp. Perhaps luminosity evolution is very strong for radio-loud quasars and did not involve normal ellipticals. It will be most interesting if optical (or other) observations show strong evidence for a massive black hole in an otherwise normal elliptical galaxy.

Our conclusions also apply to spiral galaxies if there is a similar accumulation of gas in the bulge. That this may be the case for our Galaxy is suggested by the High Energy Astrophysical Observatory HEAO-1 soft X-ray survey (see Garmire, quoted in ref. 35; J. D. Bregman, personal communication). The lack of an active nucleus within several orders of magnitude of  $L_b$  argues against the simultaneous presence of a massive black hole and a hydrostatic atmosphere in the Galactic Bulge. The recent observations<sup>36,37</sup> of high stellar velocities in the centres of M31 and M32 are also interesting in this context. There is no evidence yet of any extensive hot interstellar medium in the cores of these galaxies to which our limits can be applied. If they do contain massive black holes then mass loss from the surrounding stars may be then blown out altogether.

We have outlined a number of ways in which a black hole can reside at the centres of most ellipticals and yet be unobservable. None appears to be wholly satisfactory and if any of them proves to be correct then it is interesting in its own right. The simplest conclusion is that most bright elliptical galaxies do not contain massive black holes. The galaxies which are active now were the quasars and powerful sources of past epochs.

We thank J. Huchra for providing his galaxy luminosity function before publication. A.C.F. thanks the Royal Society for support, and C.R.C. acknowledges partial support by NASA.

Received 18 January; accepted 27 May 1988.

1. Spinrad, H. *Publs astr. Soc. Pacific* **94**, 397-415 (1982).
2. Malkan, M. A. *Astrophys. J.* **287**, 555-565 (1984).
3. Soltan, A. *Mon. Not. R. astr. Soc.* **200**, 115-122 (1982).
4. Smith, E. P., Heckman, T. M., Bothun, G. D., Romanishin, W. & Balick, B. *Astrophys. J.* **306**, 64-89 (1986).
5. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Rev. mod. Phys.* **56**, 255-351 (1984).
6. Wandel, A. & Mushotzky, R. F. *Astrophys. J.* **306**, L61-L66 (1986).
7. Sargent, W. L. W. *et al. Astrophys. J.* **221**, 731-744 (1978).
8. Young, P. J. *et al. Astrophys. J.* **234**, 76-85 (1979).
9. Lauer, T. R. *Astrophys. J.* **292**, 104-121 (1985).
10. Nulsen, P. E. J., Stewart, G. C. & Fabian, A. C. *Mon. Not. R. astr. Soc.* **208**, 185-198 (1984).
11. Begelman, M. *Nature* **322**, 614-615 (1986).
12. Gunn, J. in *Active Galactic Nuclei* (eds Hazard, C. & Mitton, S.) 213-226 (Cambridge University Press, 1979).
13. Forman, W. *et al. Astrophys. J.* **234**, L27-L33 (1979).
14. Forman, W., Jones, C. & Tucker, W. *Astrophys. J.* **277**, 19-26 (1985).
15. Thomas, P. A., Fabian, A. C., Arnaud, K. A., Forman, W. & Jones, C. *Mon. Not. R. astr. Soc.* **222**, 655-672 (1986).
16. Bondi, H. *Mon. Not. R. astr. Soc.* **112**, 195-204 (1952).
17. Canizares, C. R., Fabbiano, G. & Trinchieri, G. *Astrophys. J.* **312**, 503-513 (1987).
18. Lea, S. M., Mushotzky, R. F. & Holt, S. S. *Astrophys. J.* **262**, 24-32 (1982).
19. Stewart, G. C., Canizares, C. R., Fabian, A. C. & Nulsen, P. E. J. *Astrophys. J.* **278**, 536-543 (1984).
20. Trinchieri, G., Fabbiano, G. & Canizares, C. R. *Astrophys. J.* **310**, 637-659 (1987).
21. Rees, M. J. *A. Rev. Astr. Astrophys.* **22**, 471-506 (1984).
22. Schmidt, M. & Green, R. *Astrophys. J.* **269**, 352-374 (1983).
23. Arnaud, K. A. *et al. Mon. Not. R. astr. Soc.* **217**, 105-114 (1985).
24. Pounds, K. A. *et al. Mon. Not. R. astr. Soc.* **224**, 443-452 (1986).
25. Robinson, A., Binette, L., Fosbury, R. A. E. & Tadhunter, C. *Mon. Not. R. astr. Soc.* **227**, 97-114 (1987).
26. Jura, M. *Astrophys. J.* **306**, 483-489 (1986).
27. Dressler, A. *Astrophys. J.* **240**, L11-L16 (1980).
28. Ozernoy, L. M. in *Proc. IAU 84* (ed. Burton, W. B.) 395-398 (Reidel, Dordrecht, 1979).
29. Gurzadyan, V. G. & Ozernoy, L. M. *Astr. Astrophys.* **86**, 315-320 (1980).
30. Rees, M. J. in *The Galactic Centre* (eds Riegler, G. & Blandford, R. D.) 166-176 (Am. Inst. Phys., 1981).
31. Geballe, T. R., Wade, R., Kriscinnas, K., Gatley, I. & Bird, M. C. *Astrophys. J.* **320**, 567-569 (1987); Ozernoy, L. M. in *The Galactic Center* (ed. Backov, D. C.) 181-183 (Am. Inst. Phys., 1987).
32. Demoulin-Ulrich, M.-H., Butcher, H. R. & Boksenberg, A. *Astrophys. J.* **285**, 527-546 (1984).
33. Shapiro, S. L. *Astrophys. J.* **180**, 531 (1973).
34. Gerhard, O. E. in *Proc. IAU 127* (ed. De Zeeuw, T.) 241-248 (Reidel, Dordrecht, 1987).
35. York, D. G. *A. Rev. Astr. Astrophys.* **20**, 221-248 (1982).
36. Kormendy, J. L. in *Proc. IAU 127* (ed. De Zeeuw, T.) 17-36 (Reidel, Dordrecht, 1987).
37. Tonry, J. L. *Astrophys. J.* **283**, L27-L30 (1984).



# Ablating dwarf model for eclipsing millisecond pulsar 1957+20

E. S. Phinney, C. R. Evans, R. D. Blandford  
& S. R. Kulkarni

Theoretical Astrophysics, California Institute of Technology,  
Pasadena, California 91125, USA

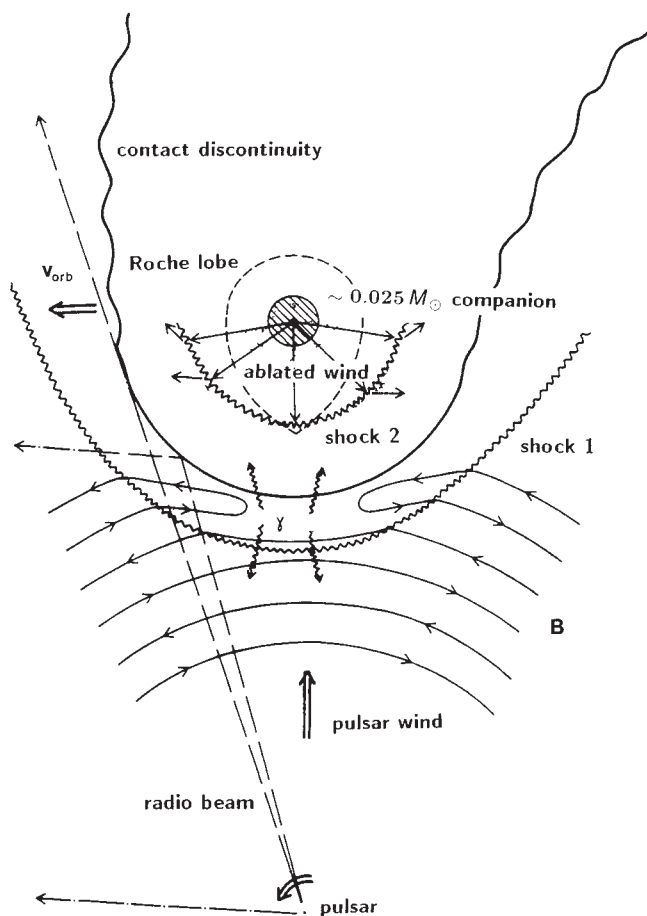
Fruchter *et al.*<sup>1</sup> have recently announced the discovery of a 1.6-millisecond pulsar, 1957+20, in a 9.2-hour eclipsing binary system and have proposed that the eclipses are due to a pulsar-excited wind derived from its  $\sim 0.024 M_{\odot}$  companion. Here we argue that the companion is a stripped, degenerate hydrogen core being ablated by the pulsar, a process that will be complete in  $\leq 10^8$  yr, leaving an isolated millisecond pulsar like 1937+21. This interpretation can be tested both by using the pulsar pulses to perform radio soundings of the interface between the two winds, and by future optical, X-ray and  $\gamma$ -ray observations. Our evolutionary model for this object predicts that there should be no hydrogen-rich, low-mass X-ray binaries (LMXB) with orbital periods  $\leq 5$  hours.

Assuming a neutron-star mass  $m_1 = 1.6 M_{\odot}$  and an inclination angle such that  $\sin i \approx 1$ , we deduce a companion mass  $m_2 \approx 0.024 M_{\odot}$  and an orbital radius  $a = 2.6 R_{\odot}$ . The Roche lobe radius<sup>2</sup>,  $R_1 \approx 0.3 R_{\odot}$  is intermediate between the inferred radius of the eclipsing region,  $R_E \approx 0.7 R_{\odot} = 5 \times 10^{10}$  cm and any plausible stellar radius; hence the wind. We argue below that the companion is most probably degenerate. Its radius is then  $R_2 \approx 0.14 \mu_e^{-5/3} R_{\odot}$ , where  $\mu_e$  is the mean molecular weight per electron<sup>3,4</sup>.

The neutron-star spin energy is  $\sim 8 \times 10^{51}$  erg and its power is  $\sim 3 \times 10^{36} \tau_8^{-1}$  erg s<sup>-1</sup>, where  $10^8 \tau_8$  yr is the pulsar timing age. (If, as we propose, the pulsar has been spun up in a mass-transfer binary, then we anticipate that  $\tau_8 \approx 1$ –10 (refs 5–7).) The companion would intercept a fraction  $\sim (R_2/2a)^2 = 8 \times 10^{-4} \mu_e^{-10/3}$ , were it not for the existence of the stand-off bow shock. A fraction  $f_w$  of this incident power can be carried off in the form of a wind, where  $f_w$  makes due allowance for the energy scattered in the form of radiation and the extra cross-section associated with the wind itself. If the wind speed  $v_w = \eta v_{\text{esc}} \approx 750(\eta/3) \mu_e^{5/6}$  km s<sup>-1</sup>, where  $v_{\text{esc}}$  is the escape speed, then the timescale for half of the degenerate dwarf to be ablated will be  $\sim 2 \times 10^6 (m_2/0.024 M_{\odot})^3 (\eta/3)^2 \mu_e^5 \tau_8 f_w^{-1}$  yr. From a detailed study of ablated winds which we will report elsewhere, we find  $f_w \approx 0.1$  and  $\eta \approx 3$  for conditions anticipated in PSR1957+20. A core derived from a young main-sequence star with cosmic abundance ( $\mu_e \approx 1.2$ ) will be degenerate for  $m_2 < 0.08 M_{\odot}$  (ref. 8) and ablation will take longest at this mass. There will be enough pulsar energy to strip the star provided  $f_w \geq 0.1$ .

The stellar wind ends at a contact discontinuity, where radio waves are eclipsed (see Fig. 1, which also describes the shock structure). Momentum flux balance with the pulsar wind determines the radius of this discontinuity,  $R_E \approx (f_w c / v_w)^{1/2} R_2$ , consistent with the observations. The eclipses appear quite sharp, with excess time delays of  $\sim 50 \mu\text{s}$  and  $\sim 400 \mu\text{s}$  near ingress and egress, respectively<sup>1</sup>. These delays suggest mean electron densities within the orbit of  $\sim 10^5 \text{ cm}^{-3}$  and  $10^6 \text{ cm}^{-3}$ , well below the stellar wind density, seemingly verifying that the contact discontinuity is quite sharp (although the thickness should be at least the Larmor radius,  $\sim 10^{7.5}$  cm, of the radiating TeV particles). The asymmetry is as expected: the orbital velocity is  $\sim 0.3 v_w$ , and the 'cometary' tail will thus be significantly curved. Polarization measurements may be able to set even more sensitive limits on the density of electrons infiltrating the pulsar wind.

The estimated field strength in the pulsar wind is  $\sim 100 \tau_8^{-1/2}$  G. Relativistic electrons and positrons of energy  $\geq 1$  TeV will be



**Fig. 1** The PSR1957+20 system, drawn to scale except for the exaggerated spacing between magnetic field reversals in the relativistic wind from the pulsar. This wind is deflected at shock 1. Particles heated at the shock and in the reconnection region cool in the  $\sim 100$  G magnetic field by emitting synchrotron and inverse Compton X-rays and  $\gamma$ -rays. These create a corona above the degenerate dwarf, and power a wind which is deflected at shock 2. The shocked pulsar wind is separated from the shocked ablated wind by a contact discontinuity where the electron density increases abruptly to  $\sim 5 \times 10^{10} \tau_8^{-1} \text{ cm}^{-3}$ . The post-shock plasma frequency exceeds the observing frequency, so the pulsar's radio beam will be reflected, and the pulsar therefore eclipsed. The duration of the intensity eclipse should be independent of frequency below  $\sim 3$  GHz. This model also predicts that as eclipse is approached an increasingly faint reflected ray will be present, delayed by  $((t - t_{\text{eci}})/2 \text{ min})^2$  pulse periods. The trailing wake caused by orbital motion makes ingress sharper than egress. At high radio frequencies near normal incidence, the beam can propagate some way into the wind before being refracted out, but is subject to free-free absorption and induced Compton losses in the plasma.

able to cool by synchrotron emission and inverse Compton scattering in the shocked pulsar wind and emit X-rays and  $\gamma$ -rays which can irradiate the surface of the degenerate dwarf. We find that soft X-rays are photoelectrically absorbed in a hot corona which drives a wind. Hard X-rays are reflected by Thomson scattering and give relatively little energy deposition, while  $\gamma$ -rays penetrate the photosphere and are re-radiated at optical wavelengths. Under the inferred conditions, the coronal density and temperature are found to be  $\sim 10^{12} \text{ cm}^{-3}$  and  $10^7$  K respectively. The density in the corona self-consistently adjusts so that heating balances cooling in the acceleration region. For X-ray

fluxes  $\gg 10^{12}$  erg cm $^{-2}$  s $^{-1}$ , cooling at the base of the wind inhibits significant increases in the mass loss.

Our interpretation of the eclipse and our estimate of the mass flux in the ablated wind imply that PSR1957+20 will ultimately destroy its companion. If the pulsar had initial spin period  $P_{\text{ms}}$  ms, it cannot have ablated a degenerate companion to its current state if the latter's initial mass exceeded  $m_{2,a} = 0.12 \mu_e^{-5/3} f_W^{1/3} P_{\text{ms}}^{-2/3} M_\odot$ . Gravitational radiation instabilities limit  $P_{\text{ms}}$  to values  $\geq 1$  (ref. 9). Millisecond pulsars are believed to be spun up by accretion, $^{5-7}$ , so the initial mass of the companion must have exceeded  $m_{2,a}$ . Because a degenerate dwarf of mass  $> m_{2,a}$  would neither have filled its Roche lobe nor been ablated, we conclude that the dwarf must have cooled from a non-degenerate state after its mass had been reduced below  $m_{2,a}$ . As  $m_{2,a}$  is of the order of the minimum mass for hydrogen burning $^8$ , this is likely to have occurred.

For accretion to spin up the neutron star, the companion must once have filled its Roche lobe. Tracing backwards from the present orbital parameters, assuming mass lost carries the specific angular momentum of the secondary (as appropriate for ablation), we find that the mass of a main-sequence secondary when it lost contact with the Roche lobe was  $m_2 = 0.6 M_\odot$ , with (assuming  $m_1 = 1.6 M_\odot$ )  $a = 1.8 R_\odot$ ,  $P_{\text{orb}} = 5$  h. A minimum of  $\sim 0.2 M_\odot$  must have been transferred to the neutron star to spin it up, so that when the secondary first contacted its Roche lobe,  $m_1 = 1.4 M_\odot$ ,  $m_2 \geq 0.8 M_\odot$ ,  $a \geq 2.2 R_\odot$ , and the orbital period was  $P_{\text{orb}} \approx 6$  h. Many LMXB have similar parameters $^{10}$ . The secondary (whose low mass allows a hydrogen core even if the system is old $^{11}$ ) must have been brought into contact with its Roche lobe by angular momentum loss, for example by gravitational radiation (effective for  $P_{\text{orb}} < 12$  h) or magnetic braking $^{12}$ .

These arguments indicate that steady mass transfer driven by angular momentum loss must have ceased when the companion's mass was reduced to  $\sim 0.6 M_\odot$ . At this mass there is a very abrupt decrease in the adiabatic stellar index ( $\zeta_{\text{ad}} = d \ln R_2 / d \ln m_2$  at a fixed specific entropy profile) from  $\sim 2.0$  to  $\sim -0.3$  as the depth of the convective envelope rapidly grows with decreasing mass $^{13}$ . Provided mass loss from the companion does not conserve angular momentum, the Roche lobe index ( $\zeta_L = d \ln R_L / d \ln m_2$ ) may become greater than  $\zeta_{\text{ad}}$ . The radius of the increasingly convective companion cannot stay within its Roche lobe and a burst of mass loss occurs. Although rapid, the mass loss will occur over many orbits, so  $a \propto (m_1 + m_2)^{-1}$ . Non-conservative mass loss may arise from early stages of pulsar ablation, X-ray ablation, or other effects $^{14,15}$ . This rapid mass loss will end when enough of the convective envelope has been lost so that as the remaining (largely radiative) star is ablated by the pulsar wind, it rapidly shrinks within its Roche lobe; accretion ends, and the orbital period (minimum  $\sim 5$  h) begins to increase. The timescale for ablation increases as the secondary shrinks.

When only  $0.085 M_\odot$  remain, nuclear burning will no longer support the star. Whether it will then cool and shrink to a degenerate state depends on the radiation flux  $F_h$  incident on the photosphere. This produces an isothermal layer of temperature  $T_i = (F_h / \sigma)^{1/4}$  and depth  $h_i$  below the surface, reducing the interior cooling flux to  $F_h / (\kappa \rho h_i)$ . Determining  $h_i$  by matching to the entropy of the main-sequence interior, we find that for  $T_i \leq 7,500$  K ( $F_h < F_c = 10^{11.3}$  erg cm $^{-2}$  s $^{-1}$ ), the cooling time is less than the ablation time, and the companion will cool to a degenerate state. For higher incident fluxes the cooling time increases so rapidly (due to the steep temperature dependence of the opacity  $\kappa_i$ ) that the companion would never become degenerate. The critical  $F_c$  corresponds to a fraction  $0.03 \tau_8$  of the flux in the pulsar wind, so the companion might be non-degenerate. Our wind models lead us to expect that this is not the case in this system, but an optical identification will decide the issue. If and when the companion becomes degenerate, the ablation timescale reaches a maximum. Subsequent ablation

proceeds more rapidly. Our model predicts that no LMXB with unevolved secondaries ( $X_H \approx 0.7$ ) should have orbital periods  $< 5$  h.

Ablation may also be important for secondaries depleted in hydrogen. For these stars, the mass-radius relation differs from the one we have assumed, perhaps allowing shorter minimum binary periods $^{16}$ . The paucity of LMXB with  $1 \leq P_{\text{orb}} \leq 3$  h (ref. 10) may be due to inhibition of mass transfer by ablation. Furthermore, an ablating secondary with  $X_H \ll 1$  may continue nuclear burning until  $m_2 \leq 0.04 M_\odot$  (refs 16, 17). This line of evolution deserves further investigation, as it might also produce isolated millisecond pulsars through binary systems similar to PSR1957+20, but with even shorter orbital periods.

The accuracy with which arrival times can be measured in PSR1957+20 may be reduced by variable dispersion due to plasma evaporated from the contact discontinuity, but multi-frequency measurements can be used to remove this. The current upper limit on the orbital eccentricity is  $e < 10^{-3}$ . A much smaller eccentricity ( $\geq 10^{-6}$ ) will be measurable once a phase-connected solution is obtained. Should there be detectable eccentricity, the relativistic periastron advance,  $\sim 1.4^\circ \text{ yr}^{-1}$  should also be measurable, allowing the pulsar mass to be determined. There is no prospect of using timing to measure orbital period changes due to gravitational radiation ( $P_b / \dot{P}_b \approx 2 \times 10^{11}$  yr, tidal or rotational timescales are not much shorter). A larger effect is the lengthening of the period due to ablation of the companion. If the timing residuals are similar to those of PSR1913+16, this should be measurable in less than ten years ( $P_b / \dot{P}_b \approx 30 \tau_{\text{abl}}$ , where  $\tau_{\text{abl}} = -m_2 / \dot{m}_2 \leq 10^9$  yr), and will determine the rate of mass loss  $-\dot{m}_2 = (m_1 + m_2)(\dot{P}_b / 2P_b)$ . Given the distance,  $\sim 1$  kpc, it should prove possible to measure a proper motion (if  $v_\perp > 10 \text{ km s}^{-1}$ ) and parallax distance.

The companion should be optically variable (presumably brightest at orbital phase 0.75) and identifiable once an accurate radio position is available. If a fraction  $f_c$  of the incident pulsar luminosity is re-radiated as optical emission, then the apparent bolometric magnitude is  $16 - 2.5 \log(f_c / \tau_8)$ . The magnitude and colour temperature will determine the radius of the star (up to uncertainties in distance and reddening), potentially discriminating between our proposed degenerate H companion, and smaller degenerate He or larger non-degenerate states. Our estimate of the effect of external heating on the cooling of the companion predicts that a degenerate object should have  $m_{\text{bol}} > 17.5$  and be cooler than 7,500 K. Once  $\tau_8$  is measured, the companion's brightness will determine the fraction of the pulsar wind luminosity radiated in forms penetrating enough to reach the photosphere ( $\kappa \leq 0.2 \text{ cm}^2 \text{ g}^{-1}$ ), a significant constraint on pulsar wind models. There is the exciting possibility that the wind speed, the mass ratio and inclination can be determined by spectroscopy. As the pulsar may be heavy from accretion, measuring its mass may put limits on the nuclear equation of state. Furthermore, it should be possible to verify our prediction that the companion has a large hydrogen fraction.

If, as discussed above, a significant fraction  $f_X$  of the energy in the shocked pulsar wind is re-radiated as  $\gamma$ -rays, the X-ray energy flux at Earth is  $\sim 10^{-9.5} f_X \tau_8^{-1}$  erg cm $^{-2}$  s $^{-1}$ , which may be measurable. Similarly, a significant fraction  $f_\gamma$  may also be converted into synchrotron or inverse Compton  $\gamma$ -rays with energy  $\sim 1$  MeV. The flux at the Earth,  $\sim 10^{-3.5} f_\gamma \tau_8^{-1}$  cm $^{-2}$  s $^{-1}$ , might be detectable by balloon experiments.

Pulsar wind ablation (a process anticipated by Ruderman, Shaham and Tavani, preprint, Columbia University) should now be considered as a mechanism for producing single millisecond radio pulsars like PSR1937+21 and the single pulsars in globular clusters, where binaries are difficult to dissociate by stellar encounters $^{18}$ . Indeed, the high rate of mass transfer induced by ablation by a pulsar wind or X-ray irradiation, and the concomitantly shortened X-ray lifetimes may ameliorate the problem of the discrepant birthrates of such pulsars and LMXB $^{19}$ . We suggest that LMXB with  $3 \leq P_{\text{orb}} \leq 12$  h ultimately produce iso-



lated millisecond pulsars as the resurrected pulsar ablates the secondary. Furthermore, as the neutron star spins up, millisecond radio pulsations might be detectable in the off states of LMXB.

We thank A. S. Fruchter and D. R. Stinebring for informing us of their exciting discovery, and D. Stevenson for useful comments. R.D.B. and C.R.E. acknowledge support from NSF; E.S.P. is a Presidential Young Investigator, and is also supported by the Irvine Foundation.

Received 10 April; accepted 24 May 1988.

1. Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. *Nature* **333**, 237-239 (1988).
2. Pennington, R. in *Interacting Binary Stars* (eds Pringle, J. E. & Wade, R. A.) 197-199 (Cambridge University Press, 1985).
3. Nelson, L. A., Rappaport, S. A. & Joss, P. C. *Astrophys. J.* **311**, 226-240 (1986).

4. Stevenson, D. J. in *Astrophysics of Brown Dwarfs* (ed. Kafatos, M. C.) 218-232 (Cambridge University Press, 1986).
5. van den Heuvel, E. P. J., van Paradijs, J. & Taam, R. E. *Nature* **322**, 153-155 (1986).
6. Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. *Nature* **300**, 728-730 (1982).
7. White, N. E. & Stella, L. *Nature* **332**, 416-418 (1988).
8. D'Antona, F. & Mazzitelli, I. *Astrophys. J.* **296**, 502-513 (1985).
9. Wagoner, R. V. *Astrophys. J.* **278**, 345-348 (1984).
10. White, N. E. in *The Evolution of Galactic X-ray Binaries* (eds Trümper, J., Lewin, W. H. G. & Brinkman, W.) 227-237 (Reidel, Dordrecht, 1986).
11. Tutukov, A. V., Fedorova, A. V., Ergma, E. V. & Yungel'son, L. R. *Astrofizika* **24**, 85-96 (1986).
12. Verbunt, F., Rappaport, S. A. & Joss, P. C. in *Cataclysmic Variables and Low-Mass X-ray Binaries* (eds Lamb, D. Q. & Patterson, J.) 39-43 (Reidel, Dordrecht, 1985).
13. Webbink, R. F. in *Interacting Binary Stars* (eds Pringle, J. E. & Wade, R. A.) 39-70 (Cambridge University Press, 1985).
14. Priedhorsky, W. C. & Verbunt, F. *Astrophys. J.* (submitted).
15. Ruderman, M. A. & Shaham, J. *Astrophys. J.* **289**, 244-246 (1985).
16. Rappaport, S. A. & Joss, P. C. *Astrophys. J.* **283**, 232-240 (1984).
17. Nelson, L. A., Rappaport, S. A. & Joss, P. C. *Astrophys. J.* **304**, 231-240 (1986).
18. Romani, R. W., Kulkarni, S. R. & Blandford, R. D. *Nature* **329**, 309-310 (1988).
19. Kulkarni, S. R. & Narayan, R. *Astrophys. J.* (in the press).

## Has comet Halley retained its primordial angular momentum?

Ignacio Ferrin\*

Grupo de Astrofísica, Departamento de Física, Universidad de los Andes, Mérida, Venezuela

In 1967, Hartmann & Larson<sup>1</sup> and Fish<sup>2</sup> pointed out a relationship between the 'angular momentum density' (angular momentum per unit mass) and mass of the planets and asteroids, which was interpreted as resulting from the primordial accretion of these objects. These authors noted that this relationship holds if the planets rotate with approximately the same period, without taking into account density variations, and Alfvén<sup>3</sup> has provided a possible mechanism for this occurrence. Here I show that the nuclei of several comets, including Halley, obey the same relationship, suggesting that these objects too retain their primordial angular momentum. The angular momentum versus mass relation depends on the nuclear dimensions, dynamical characteristics and density; if primordial accretion is assumed, knowing two of these parameters allows the determination of the third. Applying this method to comet Halley yields a bulk density of  $0.30^{+0.22}_{-0.13} \text{ g cm}^{-3}$ .

The relationship found by Hartmann and Larson is shown in Fig. 1, and represents an improvement on the relationship presented by MacDonald<sup>4</sup>. It shows the position of the planets, asteroids and comets (not included in the original diagram). Note the extraordinary character of this linear log-log relationship that extended for 11.5 orders of magnitude, and that has been now extended to 16 orders of magnitude, in mass. We will call the line that connects all the points the 'primordial line'.

To the present, comets have not been plotted on this diagram, and thus it is of interest to find out if they satisfy this relationship.

A nice opportunity has been provided by the recent observations of comet Halley<sup>5,6</sup>, the dimensions and period of rotation of which are known. Its global density has been estimated by several authors<sup>7-10</sup> (Table 1). The mean value of these determinations is  $0.28^{+0.26}_{-0.18} \text{ g cm}^{-3}$ , but some of those values are just educated guesses.

Table 2 gives the dimensions obtained for the nucleus of comet Halley by several authors<sup>11-14</sup>. The mean value of these diameters is  $15.0 \times 8.7 \times 8.8 \text{ km}$ . Thus we know the volume and the density, which allows us to calculate the total mass of the nucleus,  $1.9^{+2.5}_{-1.5} \times 10^{17} \text{ g}$ .

With this information it is possible to obtain the angular momentum of comet Halley, per unit mass. The result for this and other comets appears in Table 3. It should be noted that there have been two periods of rotation determined for comet

Table 1 Estimated densities of comet Halley ( $\text{g cm}^{-3}$ )

| Density                       | Reference |
|-------------------------------|-----------|
| 0.25                          | 8         |
| 0.1-0.3                       | 9         |
| 0.1-0.54                      | 7         |
| 0.35                          | 10        |
| Mean = $0.28^{+0.26}_{-0.18}$ |           |
| $0.30^{+0.22}_{-0.13}$        | This work |

Table 2 Dimensions\* of comet Halley

|                                                          |    |
|----------------------------------------------------------|----|
| $(14 \pm 1) \times (7.5 \pm 1) \times (7.5 \pm 1)$       | 11 |
| $14.9 \times 8.2 \times ?$                               | 12 |
| $16 \times 9 \times 10$                                  | 13 |
| $15 \times 10 \times ?$                                  | 6  |
| Adopted dimensions                                       |    |
| $(15 \pm 0.8) \times (8.7 \pm 1.1) \times (8.8 \pm 1.3)$ |    |

\* All measurements are diameters in km.

Halley, of  $54.125 \pm 0.03$  and  $177.8 \text{ h}$  (refs 14, 15). These have been interpreted by Sekanina<sup>16,17</sup> as rotation along the long axis (7.4 days), and precession along an axis fixed in space (2.2 days). Sekanina<sup>17</sup> proposed that this relatively unstable configuration is the result of the comet's nuclear splitting or of a major impact event, perhaps a collision with a relatively massive object in the not too distant past.

The angular momentum associated with the long period is much smaller than the other one, and thus does not affect our determination. Other models proposed in the literature do not affect seriously the absolute value we found for the angular momentum.

We have placed the angular momentum per unit mass for comet Halley in the  $L/M$  vs  $M$  diagram, and found that it fits the relationship quite well, falling on the primordial line. We may conclude therefore, that comet Halley is probably the result of primordial accretion.

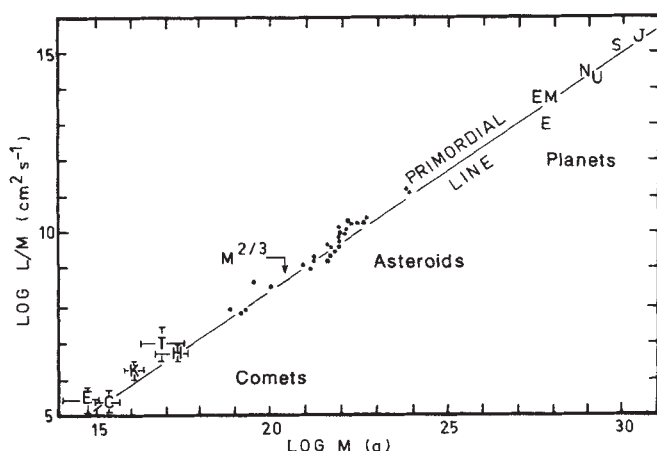
According to this result, the collision proposed by Sekanina<sup>17</sup> for the nucleus of comet Halley probably did not take place, or at most was of small intensity.

The argument can be inverted to obtain the global nuclear density. Notice that the  $L/M$  vs  $M$  relationship involves only three parameters: the nuclear dimensions (semi-axes  $a$ ,  $b$ ,  $c$ ) the global density of the nucleus,  $\rho$ , and the period of rotation,  $P_{\text{rot}}$ . If we know two of them, then we can determine the other assuming that the object lies on the primordial line. The vertical axis,  $L/M$ , is such that

$$L/M = 2\pi I / (MP_{\text{rot}}) = 2\pi(a^2 + b^2) / (5P_{\text{rot}}) \quad (1)$$

where  $a$  and  $b = c$  are the major and minor axes of the nuclear

\* Address for correspondence: Apartado 700, Mérida 5101-A, Venezuela.



**Fig. 1** The relation  $L/M$  vs  $M$  for planets, asteroids and comets. This diagram is an extended version of those presented earlier by MacDonald<sup>4</sup>, Fish<sup>1</sup> and Hartmann & Larson<sup>2</sup>. Here for the first time comets are included. Notice how well comets fit the 'primordial line'. Planets are designated by their initials (ex. EM = Earth-Moon system), asteroids by points, and comets by letters: Halley (H), Comas Sola (C), Tempel 2 (T), Kopff (K), and Encke (E). Comet Encke extends the relationship more than four orders of magnitude below the asteroids. The slope of the line is 2/3, and its equation  $\log L/M = -4.80 + 0.66 \log M$ .

**Table 3**  $L/M$  = angular momentum/mass

| Object     | $L/M$                       | Mass                               | Reference | Notes                                                                                         |
|------------|-----------------------------|------------------------------------|-----------|-----------------------------------------------------------------------------------------------|
| Halley     | $(5 \pm 1) \times 10^6$     | $1.9^{+2.5}_{-1.5} \times 10^{17}$ | This work |                                                                                               |
| Comas-Sola | $1.9 \times 10^5$           | $2.0 \times 10^{15}$               | 20        | $\rho = 1.0$ , solution I                                                                     |
| Comas-Sola | $1.7 \times 10^5$           | $2.7 \times 10^{15}$               |           | solution II                                                                                   |
| Kopff      | $1.6 \times 10^6$           | $1.1 \times 10^{16}$               | 21        | $\rho = 1.0$ , $P = 9.4$ h, $r_N = 1.45$ km                                                   |
| Kopff      | $2.7 \times 10^6$           | $1.1 \times 10^{16}$               |           | $\rho = 1.0$ , $P = 16$ h, $r_N = 2.5$ km                                                     |
| Tempel 2   | $9.5^{+9}_{-4} \times 10^6$ | $6.6^{+12}_{-4} \times 10^{16}$    | This work | Calculated using $P_{rot}$ in ref. 22, and photometric data compiled in ref. 23. See ref. 24. |
| Encke      | $(3 \pm 2) \times 10^5$     | $(6 \pm 5) \times 10^{14}$         | 25        |                                                                                               |

Units:  $L/M$ ,  $\text{cm}^2 \text{s}^{-1}$ ;  $\rho$ ,  $\text{g cm}^{-3}$ ;  $M$ , g.

ellipsoid. Thus  $L/M$  depends on the nuclear dimensions squared and the period of rotation, but not on the density.

The horizontal axis is the mass,  $M = \rho 4\pi(abc)/3$ , which depends on the density and the product of the semi-axes.

For comet Halley, the best known parameters are the period of rotation and the size (Table 2). We may obtain the density assuming the primordial hypothesis. The result is  $\rho = 0.30^{+0.22}_{-0.13}$ , not very different from the values found by other authors (Table 1). The error in  $\rho$  has been calculated from the errors in the dimensions and period of rotation, and the uncertainty involved in the primordial line.

Our determination is an independent method that has not been applied before, and it is capable of giving a rather accurate result, whereas some of the values listed in Table 1 are just educated guesses. Notice also that this result extends the  $L/M$  vs  $M$  relationship two orders of magnitude below the asteroids, by mass.

Whether other comets also lie on the primordial line, was investigated. We have located several other comets on this diagram, taken from Table 3, and found that they also satisfy it. In all cases the error bars touch the primordial line. Notice that the point of comet Encke extends the  $L/M$  vs  $M$  relationship four orders of magnitude below the asteroids.

In conclusion, we have included comet Halley and several other comets for the first time in the  $L/M$  vs  $M$  diagram, established for planets and asteroids, and have found that they

satisfy the primordial line quite well. So it can be concluded that this class of objects may be the result of primordial accretion, and that outgassing and precession, have not affected their global angular momentum by much (may be a maximum value of a factor of 2, in the present accuracy of our result). Certainly we can exclude changes by orders of magnitude, that would immediately be evident in Fig. 1. Such evidence may be found in the future.

Comets sublime, and thus the gas and dust should carry angular momentum. But because Fig. 1 involves the ratio  $L/M$ , this value should be reasonably constant with time. Also, precession does not affect the absolute value of the angular momentum, only its direction.

The fact that these comets can be placed on the primordial line, means that their rotational histories have not been changed substantially in time. This may be in contradiction with the theory of 'spin-up' of Whipple<sup>18</sup>, and with the theory of 'slow-down' of Wallis<sup>19</sup>, if their predictions are large enough to be detected at the present level of sensitivity. But none of these authors has quantified their predictions.

I thank Dr K. Wilhelm, Consejo Nacional de Investigaciones Científicas y Tecnológicas, CONICIT, Caracas and Consejo de Desarrollo Científico, Humanístico y Tecnológico, CDCHT, Universidad de los Andes, ULA, Mérida. I also thank the Fundación Polar of Caracas and the IBM Scientific Center, Caracas.

Received 25 January 1988; accepted 11 April 1988.

1. Fish, F. F. *Icarus* **7**, 251-256 (1967).
2. Hartmann, W. K. & Larson, S. M. *Icarus* **7**, 257-260 (1967).
3. Alfvén, H. *Icarus* **3**, 57-62 (1964).
4. MacDonald, G. J. F. *Space Sci. Rev.* **2**, 467-541 (1963).
5. Sagdeev, R. Z. *et al. Nature* **321**, 259-262 (1986).
6. Keller, H. U. *et al. Nature* **321**, 320-326 (1986).
7. Bertaux, J. & Abergel, A. *Eur. Space Ag. Spec. Publ.* **250**, Vol. II, 341-345 (1986).
8. Greenberg, J. M. & Grim, R. *Eur. Space Ag. Spec. Publ.* **250**, Vol. II, 255-263 (1986).
9. Rickman, H. Uppsala Preprint No. 8, Astron. Observatory, Uppsala, Sweden (1986).
10. Krasnopolsky, V. *et al. Nature* **321**, 269-271 (1986).
11. Sagdeev, R. *et al. Eur. Space Ag. Spec. Publ.* **250**, Vol. II, 307-316 (1986); *Nature* **321**, 266-268 (1986).
12. Reitsema, H. *et al. Eur. Space Ag. Spec. Publ.* **250**, Vol. II, 351-354 (1986).
13. Wilhelm, K. *et al. Eur. Space Ag. Spec. Publ.* **250**, Vol. II, 367-369 (1986).
14. Millis, R. & Schleicher, D. *Nature*, **324**, 646-649 (1986).
15. Belton, M. J. S., Wehinger, P., Wyckoff, S. & Spinrad, H. *Eur. Space Ag. Spec. Publ.* **250**, Vol. I, 599-603 (1986).
16. Sekanina, Z. In *Comet Halley 86: Worldwide Investigations, Results and Interpretations* (Ellis Horwood Library of Space, London, 1988).
17. Sekanina, Z. *Nature* **325**, 326-328 (1987).
18. Whipple, F. in *Comets* (ed. Wilkening, L.) 227-250 (University of Arizona Press, Tucson, 1983).
19. Wallis, M. *Phil. Trans. R. Soc. Lond.* **A313**, 165-170 (1984).
20. Sekanina, Z. *Astron. J.* **90**, 1370-1381 (1985).
21. Sekanina, Z. *Astron. J.* **89**, 1573-1586 (1984).
22. Barker, E. S., Cochran, A. L. & Rybski, P. M. in *Modern Observational Techniques For Comets* (ed. Brand, J. C.) 81-86, 150-155 (JPL Publication, 1981).
23. Ferrin, I. & Guzman, E. *Rev. Mexicana Astron. Astrof.* **6**, 339-345 (1981).
24. Ferrin, I. *Internal Report*, Univ. de los Andes, Dpto. de Física, Mérida, Venezuela, 1-251 (1988).
25. Sekanina, Z. *Astron. J.* **91**, 422-431 (1986).



# Synthesis, structure and superconductivity in the $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$ system

D. G. Hinks, B. Dabrowski, J. D. Jorgensen, A. W. Mitchell, D. R. Richards, Shiyong Pei & Donglu Shi

Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

The recent discovery of superconductivity near 30 K in  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  ( $x \approx 0.4$ )<sup>1,2</sup> is remarkable for two reasons. It is the first copper-free oxide superconductor that has a transition temperature ( $T_c$ ) above that for the best intermetallic superconductor; and the structure is reported to be cubic, which excludes a two-dimensional metal-oxygen sublattice analogous to the  $\text{CuO}_2$  planes believed to be responsible for superconductivity in the copper-oxide-based superconductors. Cava *et al.*<sup>1</sup> described a synthesis technique which involved starting with a 100% excess of  $\text{KO}_2$ . At least part of the excess potassium was found to be present in the final sample (in a form not detectable by X-ray diffraction), resulting in samples that were not suitable for resistivity measurements and making a precise determination of the potassium and oxygen content in the superconducting phase impossible. Here we describe a two-step synthesis technique starting with a stoichiometric oxide composition, which yields single-phase samples suitable for transport measurements. Neutron powder diffraction studies of samples with varying potassium concentration show that superconductivity in  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  occurs only in a cubic perovskite phase which is stable at  $\approx 600^\circ\text{C}$  and which forms only for  $x > 0.25$ . Within this cubic phase,  $T_c$  is highest for compositions near the structural phase transition ( $x \approx 0.25$ ) and decreases with increasing  $x$ .

Our initial attempts to synthesize  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  in oxygen or air produced only mixtures of monoclinic  $\text{BaBiO}_3$  (ref. 3) and monoclinic  $\text{KBiO}_2$  (ref. 4); this, together with Cava *et al.*'s success<sup>1</sup> in synthesizing in sealed tubes, suggested that potassium might enter the compound only in an oxygen-deficient environment. For example, if the K must enter the compound without raising the average Bi charge above +4, then oxygen vacancies must be present during the high-temperature reaction; charge balance is maintained if 1/2 oxygen vacancy is formed for each K atom introduced. Based on this hypothesis, we developed a two-part synthesis technique in which a stoichiometric mixture of oxide powders is first reacted in an inert gas, to form the oxygen vacancies and allow the incorporation of K. A low-temperature oxygen anneal is then used to fill the oxygen vacancies without allowing diffusion of K from the lattice.

Stoichiometric mixtures of oxide powders were first reacted in flowing nitrogen at temperatures from 725 to 850 °C for 1 h. The maximum reaction temperature is limited by melting, but there is no evidence that one must anneal as close as possible to the melting temperature to achieve complete reactions; in fact, the best samples were reacted at 725 °C. This initial reaction produced a dark reddish-brown powder which was not superconducting. This powder was then annealed in flowing oxygen at temperatures from 475–675 °C for 30–60 min. Oxygen anneal temperatures of  $>550^\circ\text{C}$  or extended oxygen anneal times resulted in phase separation and non-superconducting samples (see below), whereas annealing at 475–550 °C resulted in a dark blue-black powder which was superconducting. Samples for neutron powder diffraction and a.c. susceptibility measurements were made as powders in batches of 2–10 g. For transport measurements, the reddish-brown powder was pressed into pellets before the final oxygen anneal.

Superconducting transition temperatures,  $T_c$ , were measured by both a.c. susceptibility and resistivity. The onset  $T_c$ s were the same for both the inductive and resistive measurements.

**Table 1** Starting composition, oxygen anneal temperature,  $T_c$ , refined structural parameters for the cubic phase, and phases present in samples of  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$

| $x$ (init.) | $T_{\text{anneal}}$ (°C) | $T_c$ (K) | $a$ (Å)    | $x$ (ref.) | $y$ (ref.) | Phases*   |
|-------------|--------------------------|-----------|------------|------------|------------|-----------|
| 0.20        | 550                      | 0         | 4.3236 (1) |            |            | B         |
| 0.33        | 675                      | 0         | 4.3186 (1) |            |            | B + D     |
| 0.33        | 550                      | 27.1      | 4.3034 (1) | 0.25 (3)   | 0.00 (2)   | A + B     |
| 0.40        | 550                      | 27.1      | 4.2932 (1) | 0.29 (3)   | 0.02 (2)   | A         |
| 0.40        | 475                      | 21.4      | 4.2873 (1) | 0.26 (4)   | 0.07 (3)   | A         |
| 0.50        | 675                      | 0         | 4.3173 (3) |            |            | B + C + D |
| 0.50        | 550                      | 22.6      | 4.2802 (1) | 0.35 (3)   | 0.04 (2)   | A + C + D |
| 0.50        | 475                      | 15.6      | 4.2745 (1) | 0.40 (3)   | 0.03 (2)   | A + C + D |

Refinements are based on model 1 described in Table 2. Numbers in parentheses are standard deviations of the last significant digit. For the non-superconducting phase of  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  for  $x < 0.25$  (phase B), Rietveld refinement gives a pseudo-cubic lattice parameter but cannot yield values for  $x$  and  $y$ .

\*Phases present are: A,  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  ( $x > 0.25$ ), cubic, superconducting; B,  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  ( $x < 0.25$ ),  $z = 4$  supercell, non-superconducting (see text); C,  $\text{BaBiO}_3$ , monoclinic<sup>3</sup>; D,  $\text{KBiO}_2$ , monoclinic<sup>4</sup>.

Typical resistive transitions were 3 °C wide; inductive transition widths varied from 3 to 6 °C in a random way. The magnitude of the change in susceptibility at  $T_c$  was compared with that measured for a reference sample of  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$ . For all the superconducting samples the diamagnetic signal was as large or larger than that for the reference sample, thus confirming bulk superconductivity. The inductively measured onset  $T_c$ s for samples of differing starting compositions and oxygen anneal temperatures are listed in Table 1.

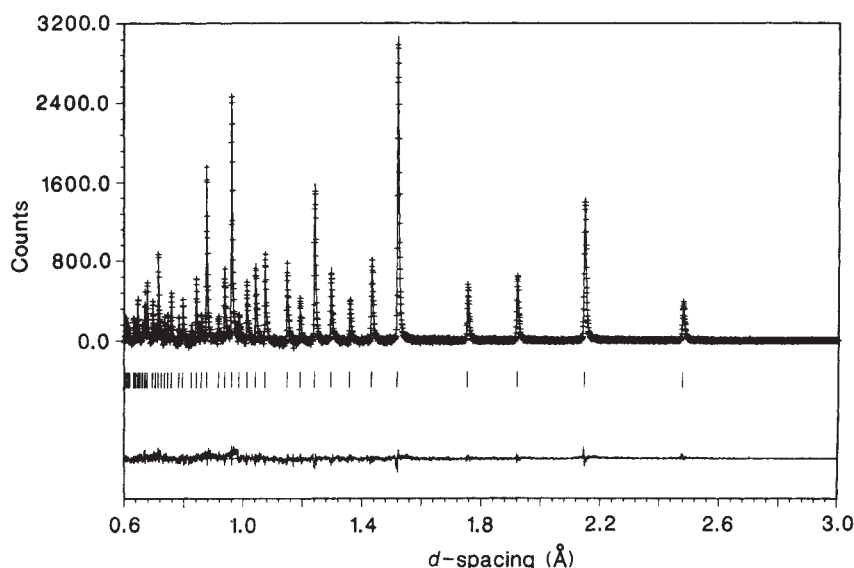
Neutron powder diffraction data were collected on the Special Environment Powder Diffractometer at the Intense Pulsed Neutron Source in runs of  $\sim 2$  h per sample. For each sample, the phases were identified and the cubic lattice parameter and chemical composition ( $x$  and  $y$ ) of the superconducting phase were determined by Rietveld refinement. The neutron diffraction results are summarized in Table 1. For samples where more than one phase was present, the minor impurity phases were included in the refinement with only their scale factors and lattice parameters being refined, to improve the precision of the refined parameters for the superconducting phase. Neutron powder diffraction was also used to examine the reddish-brown powder produced in the initial nitrogen anneal. In this way it was possible to determine if all of the initial K was present in the final superconducting phase, and, if not, whether it failed to

**Table 2** Refined structural parameters for  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  with an initial composition of  $x = 0.4$  annealed in oxygen at 550 °C

|                              | Model 1    | Model 2    |
|------------------------------|------------|------------|
| $a$ (Å)                      | 4.2932 (1) | 4.2932 (1) |
| Ba site: $n(\text{Ba})$      | 0.71 (3)   | 0.618      |
| $n(\text{K})$                | 0.29 (3)   | 0.35 (2)   |
| $n(\text{Bi})$               | 0          | 0.030 (8)  |
| $B$ (Å <sup>2</sup> )        | 1.06 (5)   | 1.06 (5)   |
| Bi site: $n(\text{Bi})$      | 1.00       | 1.000 (8)  |
| $B$ (Å <sup>2</sup> )        | 0.39 (3)   | 0.39 (3)   |
| O site: $n(\text{O})$        | 2.98 (2)   | 2.98 (2)   |
| $U_{11}^2$ (Å <sup>2</sup> ) | 0.0275 (6) | 0.0277 (6) |
| $U_{33}^2$ (Å <sup>2</sup> ) | 0.0058 (6) | 0.0059 (6) |
| $R_{\text{wp}}$              | 0.0831     | 0.0828     |
| $R_{\text{exp}}$             | 0.0682     | 0.0682     |

Refinements are done in the cubic space group  $Pm\bar{3}m$  with Ba at (0, 0, 0), Bi at ( $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$ ) and O at ( $\frac{1}{2}, \frac{1}{2}, 0$ ). Model 1 allows only Ba or K on the Ba site. Model 2 allows anti-site defects on the Ba site; that is, Ba, K, or Bi on the Ba site and no defects on the Bi site, with the Ba:Bi ratio constrained at the starting composition. Numbers in parentheses are standard deviations of the last significant digit. Parameters with no standard deviations were not refined. Refined structural parameters (site occupancies,  $n$ , isotropic temperature factors,  $B$ , components of the anisotropic temperature factor,  $U_{ii}^2$ , weighted profile  $R$  value,  $R_{\text{wp}}$  and expected  $R$  value,  $R_{\text{exp}}$ ) for  $\text{Ba}_{1-x}\text{K}_x\text{BaO}_{3-y}$ .

**Fig. 1** Rietveld refinement profile for the cubic ( $Pm\bar{3}m$ ) superconducting sample of  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  with starting composition  $x = 0.4$ , oxygen-annealed at 550 °C. Plus marks (+) are the raw time-of-flight neutron diffraction data; the solid line is the calculated profile. Tick marks below the profile mark the positions of allowed Bragg reflections included in the calculation. A difference curve (observed minus calculated) is plotted at the bottom. Background was subtracted before plotting.



react during the initial nitrogen anneal or was lost during the oxygen anneal.

Cubic, single-phase, superconducting material with  $T_c = 27.1$  K was produced for a starting composition of  $x = 0.4$  annealed in oxygen at 550 °C. Table 2 lists the complete structural parameters for a Rietveld refinement of data from this sample (Model 1); the raw data and calculated Rietveld profile are shown in Fig. 1. The refinement yields a K composition,  $x$ , of 0.29(3), which is less than the starting value, but no crystalline impurity phases are visible in the diffraction data. This suggests either an error in the initial  $\text{KO}_2$  stoichiometry (perhaps due to impurities such as carbonate or hydroxide), formation of a non-crystalline phase containing K, or the loss of some K by volatilization during the high-temperature reaction in flowing nitrogen. As shown in Table 1, this systematic discrepancy between the initial and refined K compositions is present in all of the samples for which the K concentration could be measured by neutron diffraction.

Assuming that the final sample is deficient only in K, the absence of a Bi-containing impurity phase in the diffraction data must also be explained. One possibility is that a small number of Bi atoms occupy the Ba site; such solid-solution defects are well documented in the parent compound  $\text{BaBiO}_3$  (ref. 5). This alternative structural model has also been investigated by Rietveld refinement, by constraining the overall Ba:Bi ratio at the starting composition, to investigate the magnitude of the resulting systematic error. The results of such a refinement for the sample of starting composition  $x = 0.4$ , annealed at 550 °C, are listed in Table 2 (Model 2). This alternative refinement shows that the K deficiency indicated by the neutron diffraction data is much smaller if a small number ( $\sim 3\%$ ) of Bi substitutional defects are allowed on the Ba site. From neutron diffraction alone it is impossible to tell which model is correct. As there is no corroborating evidence for the existence of anti-site defects, the following discussion will be based on the simpler model. Thus, a systematic error in the refined K concentration of the magnitude indicated ( $\sim 0.05$ ) may be present in these measurements.

From the results presented in Table 1, several important features of the structural and superconducting phase diagrams of the  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  system are evident. Superconducting samples with the cubic perovskite structure are obtained only for compositions with  $x > 0.25$ . At lower values of  $x$  there is a structural transition to a phase which is not superconducting. The sample of initial composition  $x = 0.2$ , oxygen-annealed at 550 °C is composed entirely of this non-superconducting phase. The sample with initial composition  $x = 0.33$  annealed at 550 °C contains both superconducting and non-superconducting

phases, and thus provides a fortuitous indication of the composition at the phase transition: the refined value of  $x$  is 0.25(3) (based on model 1).

For the phase with  $x < 0.25$  we observe a diffraction pattern which can be indexed on a  $z = 4$  body-centred tetragonal (or degenerate orthorhombic) cell with  $a \approx b = 6.117$  Å and  $c = 8.647$  Å ( $\sqrt{2}a \times \sqrt{2}a \times 2a$ ). This cell is similar to the  $I2/m$  monoclinic cell of  $\text{BaBiO}_3$  (which is pseudotetragonal)<sup>3</sup> or the  $I4/mcm$  tetragonal cell of superconducting  $\text{BaPb}_{1-x}\text{Bi}_x\text{O}_3$ <sup>6</sup>. However, as of this writing, no complete model has been found for this phase which agrees with the observed intensities and we have made no detailed investigation of the  $\text{BaK}_x\text{BiO}_{3-y}$  phase diagram for  $0 < x < 0.25$ .

Both samples of starting composition  $x = 0.4$  are single-phase, but the sample annealed at a lower temperature (475 °C) exhibits a lower  $T_c$ . The Rietveld refinement shows the presence of oxygen vacancies, which may explain the depressed  $T_c$ . A significant depression of  $T_c$  resulting from small oxygen vacancy concentrations has previously been reported for  $\text{BaPb}_{1-x}\text{Bi}_x\text{O}_{3-y}$ <sup>7</sup>, but an increased concentration of anti-site defects (Bi on the Ba site) might also be the cause.

Samples with starting composition  $x = 0.5$  exhibit superconductivity but are not single-phase (see Table 1), suggesting that  $x \approx 0.4$  is the solubility limit for K in this system under these synthesis conditions. This conclusion is consistent with ref. 1, in which samples of final compositions near  $x = 0.4$  were produced from starting mixtures containing a 100% excess of  $\text{KO}_2$ . Samples annealed in oxygen at 675 °C do not contain the cubic phase and are not superconducting for any starting composition, leading to the conclusion that the cubic, superconducting phase is not stable in oxygen above  $\sim 600$  °C.

The relationship between structural and superconducting behaviour is summarized in Fig. 2. As shown in Fig. 2A, the cubic lattice parameter of the superconducting phase varies monotonically with K concentration. The extrapolation of this curve to larger lattice constants provides an estimate of the K concentration of the non-superconducting samples with  $x < 0.25$ . From the refined pseudo-cubic lattice parameters for these samples (Table 1),  $x = 0.15$ – $0.18$ .

Figure 2B shows the variation of  $T_c$  with K concentration.  $T_c$  is highest adjacent to the structural transition to the non-superconducting phase at  $x = 0.25$ , and decreases with increasing  $x$ . This behaviour differs from that observed in  $\text{BaPb}_{1-x}\text{Bi}_x\text{O}_{3-y}$ , where  $T_c$  does not vary with  $x$  within the superconducting phase (near  $x = 0.25$ )<sup>8</sup>. Such behaviour suggests a possible relationship between the superconductivity and the mechanism that drives the phase transition. If the transition is driven by soft vibrational modes, the corresponding phonons



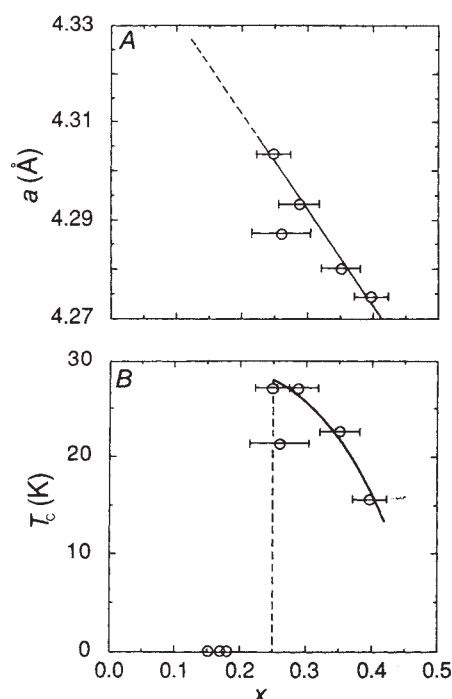


Fig. 2 Cubic lattice parameter (A) and superconducting transition temperature (B) of the superconducting phase of  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  plotted against K concentration in the cubic, superconducting phase,  $x$ , as determined by Rietveld refinement of neutron powder diffraction data. The points at  $x=0.26$  which fall below the curves are for a sample in which oxygen vacancies are present. In A, the solid line fit to the remaining points is extended to larger values of  $a$  to allow the K concentration in non-superconducting samples with  $x < 0.25$  to be estimated from the measured pseudo-cubic lattice parameter; these estimated concentrations are plotted on the  $x$ -axis in B.

may play an important role in the superconductivity. Alternatively, the transition may be electronically driven by a high density of electron states at the Fermi energy, in which case a lowering of the cell symmetry could open a gap at the Fermi energy, thereby lowering the free energy of the system. Such a 'band Jahn-Teller' transition occurs in the Chevrel-phase superconductors, giving rise to a transition from a metallic superconducting state to a semimetallic or semiconducting state which has been observed by varying pressure or composition. In the Chevrel-phase system, the highest  $T_c$ s are observed for compositions adjacent to this symmetry-lowering transition<sup>9</sup>.

The high superconducting transition temperatures observed in the  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  system, when compared with the lower  $T_c$ s (near 13 K) observed for  $\text{BaPb}_{1-x}\text{Bi}_x\text{O}_3$ , suggest that the integrity of the three-dimensional  $\text{BiO}_3$  sublattice is important for obtaining the best superconductivity in this family of compounds. Thus, future efforts to optimize the superconducting properties of the  $\text{Ba}_{1-x}\text{K}_x\text{BiO}_{3-y}$  system should concentrate on achieving a K concentration just above  $x=0.25$  and a minimum number of oxygen vacancies and anti-site defects, through careful control of the starting composition and the high-temperature nitrogen and low-temperature oxygen annealing conditions.

We thank Bruce Brown and the staff of the IPNS for providing an additional day of operation during the scheduled shut-down period expressly for these experiments. This work is supported by the US Dept of Energy, Basic Energy Sciences/Materials Sciences.

Received 1 June; accepted 14 June 1988.

1. Cava, R. J. *et al.* *Nature* **332**, 814-816 (1988).
2. Mattheiss, L. F., Gyorgy, E. M. & Johnson, D. W. Jr *Phys. Rev. B* **37**, 3745-3746 (1988).
3. Cox, D. E. & Sleight, A. W. *Solid St. Commun.* **19**, 969-973 (1976).
4. Schwedes, B. & Hoppe, R. Z. *anorg. allg. Chem.* **392**, 97-106 (1972).
5. Aurivillius, B. *Ark. Kemi Mineral. Geol.* **16A**, 1-13 (1943).

6. Cox, D. E. & Sleight, A. W. in *Proc. Conf. Neutron Scattering*, Gatlinburg, Tennessee, 45-54 (1976).
7. Suzuki, M. & Murakami, T. *Solid St. Commun.* **53**, 691-694 (1985).
8. Batlogg, B. *Physica B126*, 275-279 (1984).
9. Jorgensen, J. D., Hinks, D. G. & Felcher, G. P. *Phys. Rev. B* **35**, 5365-5368 (1987).

## Efficiency of hydrothermal ore formation and the Panasqueira W-Cu(Ag)-Sn vein deposit

David A. Polya

Department of Geology, University of Manchester, Manchester M13 9PL, UK

Hydrothermal ore bodies are the concentrated manifestations of the large-scale mass transfer of ore-forming elements by aqueous fluids flowing through the Earth's crust. Existing mass transfer estimates<sup>1-3</sup> for hydrothermal ore-forming systems are based on ore reserve data, which do not take into account the mass of ore-forming elements dispersed in the metasomatized crust around the ore bodies. Here I report a method for the quantitative determination of this dispersed mass and hence the efficiency of ore formation. Application to the Panasqueira W-Cu(Ag)-Sn vein deposit in Portugal shows that only a few per cent of the hydrothermally introduced material (W, 33%; Cu, 8%; Sn, 3%; Zn, 1.5%) is concentrated in the ore body itself, implying that much greater volumes of ore-forming fluid are required than previously supposed. The efficiency of Panasqueira vein formation in concentrating these elements is variable, reflecting the relative importance of fluid pressure<sup>4-6</sup> and acid-consuming fluid/rock reactions<sup>7</sup> in determining different ore mineral solubilities, but it is broadly similar to the efficiency of sea-water mixing in immediately removing Pb and Zn from black smoker fluids<sup>8</sup>.

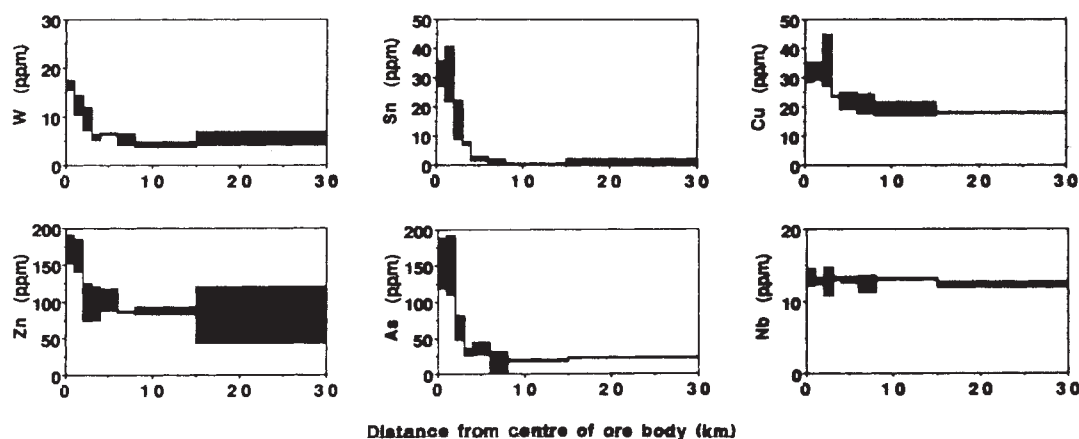
The dispersion of ore-forming elements (OFEs) by hydrothermal fluids around many vein deposits<sup>9</sup> and in the Broadlands<sup>10</sup> and East Pacific Rise (21°N)<sup>8,11</sup> systems suggests that the dispersed mass of OFEs may be particularly important in the total mass transfer balance in and around hydrothermal ore bodies. Although there are numerous studies<sup>12-18</sup> of the spatial distribution of ore-forming components around ore deposits, there is an absence of data on the total mass of dispersed components in such haloes. This reflects the difficulty in simply transforming spatial distribution data to a total mass estimate. For a simple case, where the alteration envelope is a cylinder of height  $h$  and density  $\rho$ , the total dispersed mass of a component  $i$  is given by

$$M_i = \int_0^R \rho 2\pi r h [C_i(r) - C_i^U] dr \quad (1)$$

where  $C_i(r)$  is the mean concentration (w/w) of the component in the rock at a distance  $r$  from the centre of the ore body,  $C_i^U$  is the background concentration of the component in unaltered rock, and  $R$  is the radial extent of the alteration halo, typically several metres to several kilometres. Unfortunately, for most  $C_i(r)$  functions, evaluation of the integral is particularly sensitive to changes in  $R$ , which in general cannot be determined with sufficient accuracy to meaningfully estimate  $M_i$  (see Fig. 1 for example). A method to overcome this problem is presented here and applied to a specific ore deposit.

The mass  $M_i$  of component  $i$  hydrothermally introduced into an alteration halo can be expressed by  $M_i = \rho V F^A (C_i^A - C_i^U)$ , where  $\rho$  is now the mean density of rock in the alteration halo,  $V$  its total volume,  $F^A$  the mean fraction of OFE-enriched hydrothermally altered rock in that volume, and  $C_i^A$  and  $C_i^U$  refer to the mean concentrations of component  $i$  in altered and unaltered rocks respectively.

For an isolated ore-forming hydrothermal system in a volume of isotropic crust, a normal distribution of the fraction  $F^A$  of



**Fig. 1** Variation in mean whole rock concentration of various elements in Beira Schist as a function of distance from the centre of the Panasqueira ore body. To use equation (1) to calculate the total dispersed mass of any of the components illustrated requires the determination of the radial extent,  $R$ , of the alteration halo.  $R$  is simply given by the distance at which the mean concentration of a component is equal to the background concentration. It is readily seen that  $R$  cannot be accurately determined from these data, thus necessitating the use of the method outlined in this paper. The profiles have been determined from the X-ray fluorescence analysis of 236 surface and subsurface samples. Seven samples from the small Rio deposit have been omitted. Mean values have been obtained over 1 km- to 12 km-wide annuli, as appropriate. Number of samples: 0–1 km, 29; 1–2 km, 39; 2–3 km, 21; 3–4 km, 36; 4–6 km, 30; 6–8 km, 21; 8–15 km, 33; 15–30 km, 27. Line thicknesses correspond to  $\pm 2\sigma$  for each data set.

altered rock with distance  $r$  from the ore body would be expected. All hydrothermal systems are of finite size, so  $F^A(r)$  would be expected to tend exactly to zero as  $r$  becomes large. Such behaviour may be approximated by a function, similar to the error function, and of the form

$$F^A(r) = f_0 \exp(-ar^2) \quad (2)$$

where  $f_0$  and  $a$  are empirically fitted parameters describing the magnitude and extent of the hydrothermal alteration. Thus for a common case, where the alteration envelope can be approximated by a cylinder of height  $h$  and radius  $R$ , we have

$$M_i = \int_0^\infty \rho 2\pi r h f_0 \exp(-ar^2) dr (C_i^A - C_i^U) \\ = \rho \pi h (f_0/a) (C_i^A - C_i^U) \quad (3)$$

The most important property of this equation is that, once  $(C_i^A - C_i^U)$  is determined, it is completely independent of the observed radius  $R$  of the alteration halo.

To apply these general equations to a specific deposit requires the establishment of criteria for distinguishing altered from unaltered rocks, the determination of the parameters  $f_0$ ,  $a$ ,  $h$  and  $\rho$ , and the chemical analysis of a large number of whole rock samples within the alteration halo to determine  $(C_i^A - C_i^U)$ . Application to the Panasqueira deposit is described below.

Panasqueira is a well studied<sup>17–29</sup> large vein-type W–Cu–Sn deposit associated with a post-kinematic<sup>19,21</sup> S-type<sup>17</sup> granite in northeastern Portugal. The ore body lies immediately above a greisenized cupola of the granite<sup>19</sup>. It is largely hosted by pelitic and semipelitic schists, the Beira Schists, previously metamorphosed to greenschist facies. The ore body consists largely of subhorizontal veins which, together with the scarcer subvertical and dipping veins, form a three-dimensional lacework ~100–300 metres thick and 6 km<sup>2</sup> in area. The most abundant vein mineral is quartz (88 wt%), with lesser muscovite and carbonates. The major ‘ore’ minerals are wolframite, arsenopyrite, chalcopyrite, sphalerite and cassiterite.

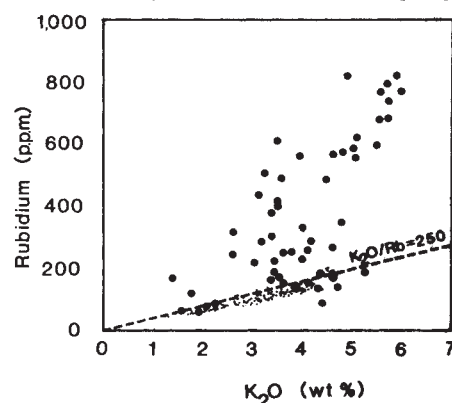
Individual veins, typically up to 1.5 metres thick are successively mantled by tourmaline-rich and muscovite-rich zones of hydrothermally altered wallrock<sup>17</sup>. The hydrothermal introduction of components such as Sn, W, Cu, As, P, Rb and F into the schists around individual veins does not extend significantly beyond the muscovite-rich zones<sup>17</sup> that extend up to several metres from the ore veins. The tourmaline-rich zones are volumetrically relatively insignificant. Hydrothermal muscovites

are Rb-rich (up to 2,500 p.p.m. Rb) compared to regional metamorphic and contact metamorphic muscovites (<400 p.p.m. Rb), hence the almost exclusive restriction of anomalous concentrations of OFEs to schist samples with whole rock  $K_2O/Rb < 250$  (Fig. 2).

Although the ore-forming hydrothermal fluids did not penetrate pervasively more than a few metres from ore veins, the occurrence of anomalous concentrations of OFEs up to 10 km from the mine indicates that the fluids moved large distances along widely spaced vein or fracture networks. Sampling ( $n = 243$ ) of the dispersion halo around Panasqueira therefore revealed considerable heterogeneity on a metre-scale.

Trends of OFE enrichment could be assessed on a kilometre-scale by determining the mean composition of the Beira Schists in a series of annuli around the centre of the Panasqueira ore body. These annuli are increasingly enriched in ore-forming components closer to the ore body (Fig. 1). Using  $K_2O/Rb < 250$  as a criterion for hydrothermal alteration of individual samples (compare Fig. 2), it was determined that this pattern is not due mainly to systematic changes in the chemistry of individual samples, but to an increasing proportion of hydrothermally altered samples in the annuli closer to the centre of the ore body (Table 1, Fig. 3).

The trend in Fig. 3 suggests that Panasqueira is the largest hydrothermal ore body within the zone of sampling. If the effect



**Fig. 2** Beira Schist whole rock  $K_2O$  versus Rb, showing the correlation between low  $K_2O/Rb$  and anomalously high whole rock Sn. High Sn (>5 p.p.m.) indicated by filled circles, low or background Sn (<5 p.p.m.) by dots (based on the X-ray fluorescence analysis of 243 samples).



**Table 1** Variation in the mean composition (in p.p.m.) of altered and unaltered schists with distance from the centre of the Panasqueira ore body

|                          | 0-1 km   | 1-2 km   | 2-4 km   | 4-8 km  | Mean     |
|--------------------------|----------|----------|----------|---------|----------|
| W Altered                | 18 ± 3   | 16 ± 3   | 20 ± 10  | 9 ± 2   | 18 ± 7   |
| W Unaltered              | 5 ± 1    | 5 ± 1    | 5 ± 3    | 6 ± 1   | 6 ± 2    |
| Sn Altered               | 36 ± 6   | 45 ± 8   | 55 ± 32  | 14 ± 6  | 46 ± 21  |
| Sn Unaltered             | 3 ± 2    | 5 ± 1    | 4 ± 1    | 1 ± 2   | 2 ± 2    |
| Cu Altered               | 33 ± 8   | 38 ± 5   | 88 ± 44  | 28 ± 9  | 57 ± 29  |
| Cu Unaltered             | 28 ± 6   | 21 ± 3   | 21 ± 2   | 21 ± 2  | 21 ± 2   |
| Zn Altered               | 181 ± 20 | 177 ± 17 | 137 ± 25 | 86 ± 31 | 158 ± 22 |
| Zn Unaltered             | 114 ± 14 | 134 ± 18 | 93 ± 4   | 91 ± 5  | 92 ± 5   |
| As Altered               | 175 ± 42 | 216 ± 42 | 212 ± 87 | 79 ± 29 | 202 ± 64 |
| As Unaltered             | 18 ± 5   | 20 ± 3   | 20 ± 2   | 26 ± 10 | 25 ± 9   |
| $M_a$ ( $10^9$ tonnes)   | 4.4      | 13.2     | 52.8     | 211.0   | 281.4    |
| $M_a^A$ ( $10^9$ tonnes) | 3.8      | 7.7      | 8.5      | 0.9     | 20.9     |
| $M_a^U$ ( $10^9$ tonnes) | 0.6      | 5.5      | 44.3     | 210.1   | 260.5    |
| $M_a^A/\Sigma M_a^A$ (%) | 18       | 37       | 41       | 4       | (100)    |
| $M_a^U/\Sigma M_a^U$ (%) | 0.2      | 2        | 17       | 81      | (100)    |

Each column shows data for an annulus of height 500 m and inner and outer radii given in the column heading.  $M_a$ ,  $M_a^A$  and  $M_a^U$  refer respectively to the total mass, mass of altered schists and mass of unaltered schists within each annulus. The mean W, Sn, Cu, Zn and As content of altered and unaltered schists within the whole alteration envelope around Panasqueira (shown in the right-hand column) has been calculated from  $C_i^A = \Sigma((C_{ia}^A)(M_a^A/\Sigma M_a^A))$  and  $C_i^U = \Sigma((C_{ia}^U)(M_a^U/\Sigma M_a^U))$  respectively. Note that well over 99% of the total mass of altered schist lies within the four annuli, 0-1 km, 1-2 km, 2-4 km and 4-8 km from the ore body. The consideration of  $C_{ia}^A$  and  $C_{ia}^U$  for annuli further from the ore body is inconsequential to the calculation of  $C_i^A$  and  $C_i^U$ .

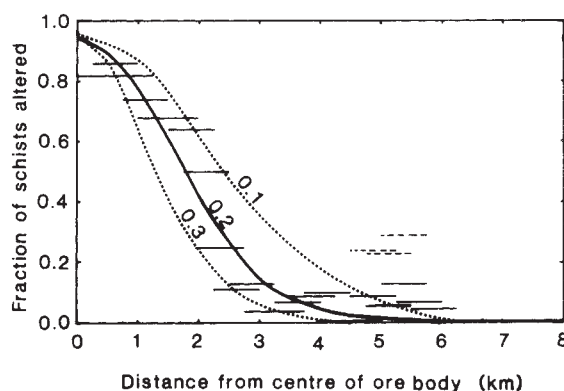
Errors reported are standard errors,  $\sigma_N/\sqrt{N}$ , where  $N$  is the number of samples analysed: 0-1 km 25 altered, 4 unaltered; 1-2 km 26, 13; 2-4 km 7, 50; 4-8 km 3, 48. Omitted from the calculations are 7 samples from the Rio prospect, as well as analyses of minor quartzites, K-rich metasediments and dolerites with the Beira Schists.

of a satellite hydrothermal system around the much smaller Rio deposit is filtered out, then  $F^A(r)$  may be approximated by a function of the form of equation (2), with an adequate fit to the analytical data being given by the selection of the values  $f_0 = 0.95$  and  $a = 0.20 \text{ km}^{-2}$ .

There appears to be no systematic difference in  $C_i^U$  for the various kilometre-scale annuli around Panasqueira (Table 1). This strongly suggests that, in this region, the Beira Schists, excluding minor (<5%) quartzites, K-rich metasediments and dolerite dykes, are essentially homogeneous on a kilometre-scale. There is no indication from this data of a significantly OFE-depleted recharge zone within 8 km of Panasqueira.

The height,  $h$ , of the alteration halo is estimated to be ~500 m: this rather conservative value lies between the 300 metre thickness of the ore body and the 0.6-2 km depth estimated for Panasqueira ore formation<sup>19,28</sup>. Beira schist densities<sup>29</sup> are around 2,800 million tonnes  $\text{km}^{-3}$ .

The masses of ore-forming components introduced into the Beira Schist by the Panasqueira hydrothermal system are thus calculated from equation (3) to be  $0.3 \times 10^6$  tonnes W,  $0.8 \times 10^6$  tonnes Cu,  $1.4 \times 10^6$  tonnes Zn,  $0.9 \times 10^6$  tonnes Sn, and  $4 \times 10^6$  tonnes As. The Panasqueira alteration halo is rather more extensive in the NW-SE direction than in the SW-NE direction, and the vertical extent of the halo decreases somewhat with increasing distance from the ore body. The errors introduced here by modelling the alteration halo more simply as a cylinder are estimated to be of the same order as the errors in the determination of  $C_i^A$  and  $C_i^U$  (Table 1). The calculated masses, estimated to be accurate to within a factor of 2, are mostly considerably greater than the masses of the same components in the Panasqueira vein system (Table 2). The vein ore body is therefore shown to be a quantitatively minor manifestation of a larger regional hydrothermal alteration system.

**Fig. 3** Fraction ( $F^A$ ), of Beira Schist that is hydrothermally altered as a function of distance ( $r$ ) from the apex of the Panasqueira cupola. Bars show running means over 0.75 km intervals. For each bar,  $F^A$  has been calculated by  $F^A = N^A/(N^A + N^U)$ , where  $N^A$  and  $N^U$  are the number of altered and unaltered samples respectively in the sample set. The dashed bars indicate the perturbation caused by the inclusion of 7 samples from the much smaller Rio deposit. Curves show the function:  $F^A(r) = 0.95 \exp(-ar^2)$  for  $a = 0.10$ ,  $a = 0.20$  and  $a = 0.30$ .

The ratio of concentrated to dispersed plus concentrated material (Table 2) may be used as an indicator of the maximum time-integrated efficiency of an ore-forming system. The high efficiency (33%) of vein dilation in concentrating W from the Panasqueira ore-forming fluids suggests a positive pressure-dependence of wolframite solubility, similar to that of quartz<sup>6</sup>. Continuous removal of Cu and Zn throughout the Beira Schists by acid-consuming fluid/wallrock<sup>7</sup> reactions, together with the generally inverse pressure-dependence of sulphide mineral solubilities<sup>4,5</sup>, accounts for the low efficiency (8% for Cu, 1.5% for Zn) of vein dilation in concentrating these elements. It is interesting to note that these efficiencies are comparable to those (3%) of immediate removal of Pb and Zn from black smoker hydrothermal fluids by sea-water mixing<sup>8</sup>.

At least  $5 \times 10^9$  tonnes of a hydrothermal fluid containing 5 p.p.m. Sn would be required to precipitate the 25,000 tonnes of Sn in the Panasqueira ore veins. But about 30-40 times this amount of fluid would be required to precipitate the 900,000 tonnes of Sn hydrothermally introduced into the dispersion halo. If the large ratio of dispersed to concentrated hydrothermal material at Panasqueira pertains to other ore deposits, then calculations of the total flux of hydrothermal fluid required for ore formation, based on ore mineral solubilities and ore reserve sizes<sup>1-3</sup>, may be underestimates by several orders of magnitude.

**Table 2** Masses of hydrothermally transported material concentrated in the Panasqueira ore body and dispersed in the surrounding alteration halo

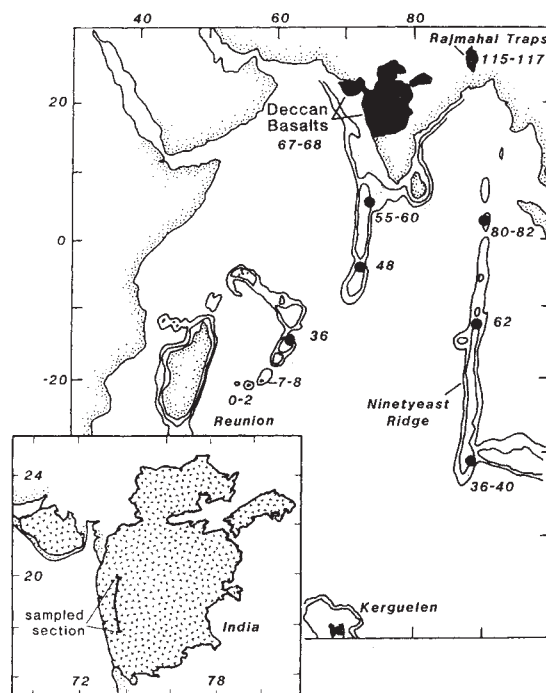
| Transported component | Ore body (tonnes) | Halo (tonnes) | $C/(D+C)$ |
|-----------------------|-------------------|---------------|-----------|
| W                     | 150,000           | 300,000       | 33%       |
| Cu                    | 70,000            | 800,000       | 8%        |
| Sn                    | 25,000            | 900,000       | 3%        |
| As                    | 100,000           | 4,000,000     | 2.5%      |
| Zn                    | 20,000            | 1,400,000     | 1.5%      |

The column headed  $C/(D+C)$  is the % of the total hydrothermally transferred mass of a given component that is concentrated within the ore body itself. Ore body calculations are based on visual estimates of ore mineral abundances relative to the abundances of wolframite, chalcopyrite and cassiterite+stannite, determined from mine production data. Estimated bulk composition of the ore veins includes 13,000 p.p.m. W, 8,000 p.p.m. As, 6,000 p.p.m. Cu, 2,000 p.p.m. Sn and 1,700 p.p.m. Zn. Ore vein mass is estimated to be about 12 million tonnes.

The author thanks the management and staff of Beralt Tin and Wolfram (Portugal) for access and logistic support. X-ray fluorescence analysis data were reduced by P. Harvey and B. Atkin. I also thank M. Solomon, R. A. D. Patrick, D. Rickard and R. P. B. Hebblethwaite for constructive criticism of the manuscript. The work was supported by the NERC and the Royal Commission for the Exhibition of 1851.

Received 29 February; accepted 17 May 1988.

1. Elder, J. W. *Geothermal Systems* (Academic, London, 1981).
2. Cathles, L. M. *Econ. Geol.* **5**, 439-487 (1983).
3. Campbell, A., Rye, D. & Petersen, U. *Econ. Geol.* **79**, 1818-1832 (1984).
4. Hemley, J. J., Cygan, G. L. & d'Angelo, W. M. *Geology* **14**, 377-379 (1986).
5. Helgeson, H. C. & Lichtner, P. C. *J. geol. Soc. Lond.* **144**, 313-326 (1987).
6. Fournier, R. O. *Rev. econ. Geol.* **2**, 45-62 (1985).
7. Eugster, H. P. *Mineralog. Mag.* **49**, 7-23 (1985).
8. Converse, D. R., Holland, H. D. & Edmond, J. M. *Earth planet. Sci. Lett.* **69**, 159-175 (1984).
9. Rose, A. W., Hawkes, H. E. & Webb, J. S. *Geochemistry in Mineral Exploration* (Academic, New York, 1979).
10. Ewers, G. R. & Keays, R. R. *Econ. Geol.* **72**, 1337-1354 (1977).
11. Edmond, J. M., von Damm, K. L., McDuff, R. E. & Measures, C. I. *Nature* **297**, 187-196 (1982).
12. Boyle, R. W. (ed.) *Geochemical exploration: Canadian Inst. Min. Metall.* Vol. 2 (1971).
13. Govett, G. J. S. & Pantazis, Th. M. *Trans. Inst. Min. Metall.* **B80**, 27-46 (1971).
14. Polya, D. A., Solomon, M., Eastoe, C. J. & Walshe, J. L. *Econ. Geol.* **81**, 1341-1355 (1986).
15. MacLean, W. H. & Kranidiotos, P. *Econ. Geol.* **82**, 951-962 (1987).
16. Lavery, J. H. *J. geochem. Explor.* **23**, 35-60 (1985).
17. Bussink, R. W. *Geologica Ultraetina* **33** (1984).
18. Oosterom, M. G., Bussink, R. W. & Vriend, S. P. *Miner. Deposita* **19**, 283-288 (1984).
19. Kelly, W. C. & Rye, R. O. *Econ. Geol.* **74**, 1721-1822 (1979).
20. Clark, A. H. *Trans. Inst. Min. Metall.* **73**, 813-824 (1964).
21. Kelly, W. C. *Servicos Geol. Portugal Commun.* **81**, 239-244 (1977).
22. d'Orey, F. C. *Servicos Geol. Portugal Commun.* **52**, 117-167 (1967).
23. Marignac, M. C. *C.R. Acad. Sci. Paris* **277D**, 269-272 (1973).
24. Marignac, M. C. *Econ. Geol.* **77**, 1263-1266 (1982).
25. Polya, D. A. *Abstr. Vol. Min. Dep. Study Group* (University of Southampton, 1986).
26. Wimmers, D. *Mineralog. Mag.* **49**, 745-748 (1985).
27. Thadeu, D. *Servicos Geol. Portugal Commun.* **32**, 5-64 (1951).
28. Bussink, R. W., Kreulen, R. & de Jong, A. F. M. *Bull. Mineralog.* **107**, 703-713 (1984).
29. Mello Mendes, F. *Comportamento Mecanico de Rochas Xistosas* (Imprensa Portuguesa, Porto, 1960).



**Fig. 1** The Deccan flood basalts lie at the northern end of the volcanic trace of the Reunion hotspot. The ages (shown in Myr) of the volcanism decrease to the south from Deccan lavas (our  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  dating of tholeiitic basalts from a section ~2,000-m-thick<sup>6</sup>, shown inset) to Ocean Drilling Program Sites 715, 713 and 706 (biostratigraphic age estimates<sup>20</sup>) to basalts from Mauritius and Reunion Islands (K-Ar dating<sup>21</sup>). A parallel companion volcanic trace starts with the Rajmahal basalts<sup>11</sup>, follows the Ninetyeast Ridge<sup>10</sup>, and ends with the youngest activity near Kerguelen Island.

## Rapid eruption of the Deccan flood basalts at the Cretaceous/Tertiary boundary

R. A. Duncan & D. G. Pyle

College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA

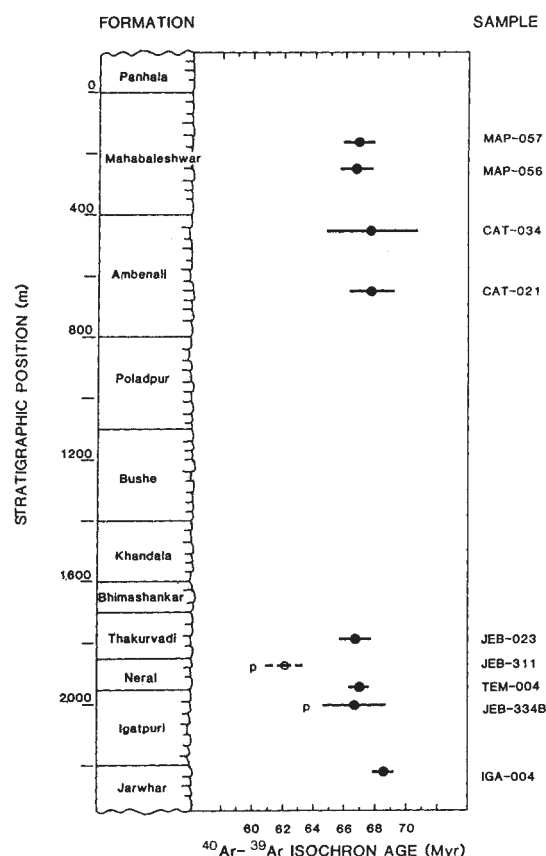
The accumulation of flood basalts of the Deccan Traps, western India, is one of the most remarkable volcanic provinces on Earth in sheer extent and volume. These rocks are akin in composition and occurrence to other extensive continental basalt provinces, usually located near plate margins, but distributed worldwide and of various ages<sup>1</sup>. Recent interest in the Deccan flood basalts has increased with the suggestion that this volcanic activity may have coincided with mass faunal extinctions and/or meteorite impact at the Cretaceous/Tertiary boundary<sup>2-5</sup>; critical to these correlations are the age and age range of the volcanic rocks. Here we report  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  geochronological results from stratigraphically controlled, geochemically well characterized tholeiitic basalts from the thickest sequence of the Deccan flood basalts, Western Ghats<sup>6</sup>. The samples dated span a section ~2,000 m thick, and range in age from 66.6 to 68.5 Myr. There is no significant difference in age from the stratigraphically oldest to the youngest rocks. It appears, then, that this sequence of lavas was erupted rapidly and can be correlated in time with events at the Cretaceous/Tertiary boundary.

The origin of flood basalt volcanism has been problematic and the plate tectonic paradigm has not offered a solution. A strong correlation has been noted, however, between such massive volcanic provinces and the initiation of hotspot activity, usually during continental breakup<sup>7</sup>. Morgan<sup>7,8</sup> proposed that

the Deccan flood basalts were the first manifestation of the stationary mantle hotspot that subsequently produced the volcanic ridge underlying the Laccadive, Maldiva and Chagos Islands, the Mascarene Plateau, and the young volcanic islands of Mauritius and Reunion (Fig. 1). The primary evidence for this hypothesis is that the geometry and northward age progression of these volcanic islands and submarine ridges is consistent with northward motion of the Indian plate, followed by northeastward motion of the African plate, over a fixed melting anomaly near the location of Reunion throughout Tertiary time<sup>9,10</sup>. In addition, this volcanic trail is parallel with the Ninetyeast Ridge, another age-progressive, submarine lineament linked to hotspot activity<sup>10</sup> (now centred near the Kerguelen Islands on the Antarctic plate). The Kerguelen hotspot also began with flood basalt volcanism, that of the Rajmahal Traps, eastern India, but much earlier (115 to 117 Myr)<sup>11</sup>.

The absolute age and age range of the Deccan basalts are thus important in the context of the hotspot model and the position of India relative to the Reunion hotspot in latest Cretaceous and early Tertiary time. Of enormous additional interest is the relationship of events at the time of the Cretaceous/Tertiary boundary. An alternative position to the meteorite impact hypothesis<sup>12</sup> is a volcanic catastrophe, which Officer and Drake<sup>2</sup> have proposed to be the Deccan flood basalt event. Indeed, a general correlation between flood basalt eruptions and mass extinctions has been made<sup>3</sup>. The associated hotspots would then be the waning tails of initially large mantle outgassing events. This second model applies, however, only if it can be shown that the volcanism coincided exactly with the Cretaceous/Tertiary boundary and occurred over a short period. Conversely, precise correlation of the Deccan basalts with the





**Fig. 2**  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  isochron ages for Deccan basalts from the main tholeiitic section, Western Ghats<sup>6</sup> are plotted against stratigraphic position of the samples (data from Table 1). Uncertainties are  $\pm 1$  s.d.; the two samples marked with a 'p' are plagioclase separates, all others are whole rock. With one exception, sample ages are indistinguishable throughout the section and give a mean age of  $67.4 \pm 0.7$  Myr. Hence, most of the Deccan volcanism occurred within perhaps one million years very near the Cretaceous/Tertiary boundary.

Cretaceous/Tertiary boundary may also link the volcanism to the site of meteorite impact<sup>4,5</sup>.

Radiometric dating of Deccan basalts has been mostly by the K-Ar method, and measured ages range from greater than 100 Myr to 30 Myr (for a summary see ref. 5). The K-Ar ages can be affected by low-temperature alteration, which may add K and remove radiogenic Ar, thus causing measured ages to be significantly less than the crystallization age. Alternatively, assimilation of older continental crust may add radiogenic Ar at the time of crystallization, causing measured ages to be too old. It appears that the least altered of Deccan basalts yield K-Ar ages between 60 and 67 Myr<sup>5,13</sup>. As all samples show some degree of secondary alteration, it is crucial to perform  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  incremental heating experiments to assess the internal consistency of age data from these rocks.

Kaneoka<sup>14</sup> reported  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  analyses on several Deccan basalts, but only two of these yielded well defined plateau ages of 68 and 63 Myr. These ages appear to be reliable and come from the Girnar Hills and Igatpuri localities respectively. Venkatesan *et al.*<sup>15</sup> reported three  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  incremental heating ages from the Mahabaleshwar region between 65 and 72 Myr, but large analytical uncertainties have led Baksi<sup>13</sup> to question the reliability of the ages.

Samples analysed in the present study come from a composite sequence ~2,000-m-thick of tholeiitic flood basalts from the Western Ghats<sup>6</sup> (Fig. 1, inset). The volcanic section, from the Jawhar to Mahabaleshwar Formations, is represented; the relative stratigraphic positions of the samples are shown in Fig. 2. In thin section the rocks vary from aphyric to porphyritic basalts, some containing remarkably large (2-3 cm) plagioclase phenocrysts. All samples show low-temperature alteration of once glassy groundmass to smectite, zeolite and calcite. Plagioclase phenocrysts are occasionally saussuritized, and less common olivine phenocrysts are iddingsitized. Hence no sample was considered ideal for dating purposes, but the seven freshest whole rock basalts and two plagioclase concentrates were selected for  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  incremental heating experiments. After crushing and ultrasonic washing in distilled water, each of the samples (~1 g) was irradiated<sup>16</sup> and argon extracted and analysed in a series of seven to eight steps of increasing temperature, the last resulting from fusion of the sample.

Age calculations from the  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  analytical data are given in Table 1 and four representative argon release patterns are

**Table 1**  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  ages for Tholeiitic basalts from the Deccan Volcanic Province, western India, from the Jawhar to Mahabaleshwar Formations

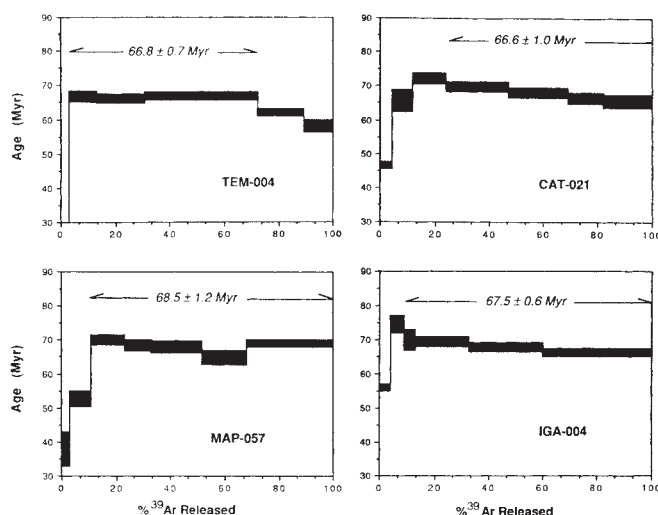
| Sample number*         | Total fusion age (Myr) | Plateau age† (Myr)<br>(% $^{39}\text{Ar}$ released) | Isochron age‡ (Myr) | $^{40}\text{Ar}$ - $^{36}\text{Ar}$ intercept |
|------------------------|------------------------|-----------------------------------------------------|---------------------|-----------------------------------------------|
| MAP-057                | $57.3 \pm 1.3$         | $68.5 \pm 1.2$<br>(89%)                             | $66.9 \pm 1.0$      | $305 \pm 11$                                  |
| MAP-056                | $61.2 \pm 0.4$         | $64.7 \pm 1.4$<br>(65%)                             | $66.7 \pm 1.0$      | $289 \pm 19$                                  |
| CAT-034                | $67.0 \pm 0.6$         | $64.4 \pm 0.9$<br>(94%)                             | $67.7 \pm 3.0$      | $286 \pm 12$                                  |
| CAT-021                | $68.1 \pm 0.6$         | $66.6 \pm 1.0$<br>(76%)                             | $67.7 \pm 1.5$      | $289 \pm 15$                                  |
| JEB-023                | $64.1 \pm 0.7$         | $67.8 \pm 1.8$<br>(67%)                             | $66.7 \pm 1.0$      | $299 \pm 3$                                   |
| JEB-311 (plagioclase)  | $59.1 \pm 0.8$         | $60.8 \pm 2.1$<br>(63%)                             | $62.1 \pm 1.2$      | $295 \pm 2$                                   |
| TEM-004                | $68.9 \pm 0.6$         | $66.8 \pm 0.7$<br>(70%)                             | $66.9 \pm 0.6$      | $295 \pm 2$                                   |
| JEB-334B (plagioclase) | $59.7 \pm 2.6$         | $62.7 \pm 2.1$<br>(60%)                             | $66.6 \pm 2.0$      | $288 \pm 4$                                   |
| IGA-004                | $67.6 \pm 0.6$         | $67.5 \pm 0.6$<br>(91%)                             | $68.5 \pm 0.6$      | $279 \pm 20$                                  |

Flux monitor MMhb-1:  $519.5 \pm 2.5$  Myr<sup>16</sup>. Decay constants used in calculations:  $\lambda_e = 0.581 \times 10^{-10} \text{ yr}^{-1}$ ,  $\lambda_\beta = 4.962 \times 10^{-10} \text{ yr}^{-1}$ .

\* From Beane *et al.*<sup>1</sup>.

† Weighted mean of ages from adjacent gas increment steps.

‡ Calculated from  $^{40}\text{Ar}$ - $^{36}\text{Ar}$  versus  $^{39}\text{Ar}$ - $^{36}\text{Ar}$  data for plateau steps.



**Fig. 3** Representative incremental heating experiments are illustrated by age versus proportion of total gas released for four samples reported in Table 1. Plateau ages are calculated from three or more contiguous gas increments that yield the same apparent age ( $1\sigma$  uncertainties are indicated). Isochron ages calculated from the same gas compositions which define the plateaux are concordant with the plateau ages.

shown in Fig. 3. Total fusion  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  ages determined from a split of each of the irradiated samples vary from 57 to 69 Myr. Plateau and isochron ages calculated from the incremental heating experiments, however, give a more restricted age range.

Sample ages reported in Table 1 satisfy the conservative selection criteria of Lanphere and Dalrymple<sup>17</sup>. The age plateau for a given sample is defined by a continuous set of incremental gas compositions which yield the same apparent age and represent a significant fraction of the  $^{39}\text{Ar}$  released during heating. The proportion of the total gas ( $\%^{39}\text{Ar}$ ) represented in the plateau steps varies from 60 to 94%. The isochron age is derived from the slope of the best fitting line to the  $^{40}\text{Ar}$ - $^{36}\text{Ar}$  versus  $^{39}\text{Ar}$ - $^{36}\text{Ar}$  data for the same gas release steps that define the plateau. The plateau and isochron ages are concordant, within analytical uncertainty ( $1\sigma$ ). The  $^{40}\text{Ar}$ - $^{36}\text{Ar}$  intercepts are not different from the atmospheric composition (296), except for plagioclase sample JEB-334B, which produced a slightly lower intercept. There is no evidence that any of the samples have been affected by inherited radiogenic argon.

With the exception of the two plagioclase separates (JEB-311 and JEB-334B), all samples yielded isochron ages between 66.7 and 68.5 Myr (corresponding plateau ages are 64.4 to 68.5 Myr). The analytical data offer no reason to discount the youngest age (JEB-311), but it is clearly anomalous considering its stratigraphic position. The gas release spectrum for sample JEB-334B is saddle-shaped and the plateau age is based on the two highest age increments only. The plateau and isochron ages are barely concordant, and the  $^{40}\text{Ar}$ - $^{36}\text{Ar}$  intercept is slightly less than the atmospheric value, so this sample age is not considered reliable. A weighted least-squares regression of age on stratigraphic position (Fig. 2) determined that there is no significant difference in age between the lowermost and uppermost samples. We conclude that this major portion of the Deccan flood basalts was erupted very rapidly.

Courtillot *et al.*<sup>5</sup> have argued that only two reversals of the geomagnetic field are recorded in the Deccan lava flows and that the entire basaltic section accumulated in perhaps one million years, predominantly during reversed polarity chron C29R (ref. 18). Our data support this catastrophic view of Deccan volcanism. The average eruption rate for this province would have been about  $1\text{ km}^3$  per year, although episodes of more rapid eruption separated by quiet periods are more likely.

The formal mean age for the sampled section is  $67.4 \pm 0.7$  Myr. This corresponds closely with the latest estimate for the age of the Cretaceous/Tertiary boundary (66.4 Myr)<sup>18</sup> and supports the proposed time correlation between Deccan volcanism and mass extinctions at that time<sup>3</sup>, with or without the connection of meteorite impact. Other similarly well dated flood basalt provinces, such as the Rajmahal Traps ( $116 \pm 1$  Myr)<sup>11</sup> and the majority of the Columbia River Basalts ( $16 \pm 1$  Myr)<sup>19</sup> were also erupted over a short period. We conclude that further investigation of the climatic effects of rapid eruption of flood basalts is warranted.

We thank P. R. Hooper for samples and helpful discussions. This work was supported by the US Geological Survey (for sample irradiations) and the National Science Foundation.

Received 14 March; accepted 23 May 1988.

1. Basaltic Volcanism Study Project *Basaltic Volcanism on the Terrestrial Planets* (Pergamon, New York, 1981).
2. Officer, C. B. & Drake, C. L. *Science* **219**, 1383-1390 (1983).
3. Rampino, M. R. & Stothers, R. B. *Trans. Am. geophys. Un.* **67**, 1247 (1986).
4. Hartnady, C. J. H. *Geology* **14**, 423-426 (1986).
5. Courtillot, V. *et al. Earth planet. Sci. Lett.* **80**, 361-374 (1986).
6. Beane, J. E., Turner, C. A., Hooper, P. R., Subbarao, K. V. & Walsh, J. N. *Bull. volcan.* **48**, 1-33 (1986).
7. Morgan, W. J. *Geol. Soc. Am. Mem.* **132**, 7-22 (1972).
8. Morgan, W. J. in *The Sea* (ed. Emiliani, C.) 443-487 (Wiley, New York, 1981).
9. Duncan, R. A. *Tectonophysics* **74**, 29-42 (1981).
10. Duncan, R. A. *J. Volcan. geotherm. Res.* **4**, 283-305 (1978).
11. Baksi, A. K. *Terra Cognita* **6**, 161 (1986).
12. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095-1108 (1980).
13. Baksi, A. K. *Geology* **15**, 147-149 (1987).
14. Kaneoka, I. *Earth planet. Sci. Lett.* **46**, 233-243 (1980).
15. Venkatesan, T. R., Pande, K. & Gopalan, K. *J. geol. Soc. India* **27**, 102-109 (1986).
16. Dalrymple, G. B., Alexander, E. C., Lanphere, M. A. & Kraker, G. P. *US Geological Survey Prof. Paper* 1176 (1981).
17. Lanphere, M. A. & Dalrymple, G. B. *US Geol. Surv. Open File Rpt* **78-701**, 241-243 (1978).
18. Berggren, W. A., Kent, D. V., Flynn, J. J. & van Couvering, J. A. *Geol. Soc. Am. Bull.* **96**, 1407-1418 (1985).
19. Long, P. E. & Duncan, R. A. *Trans. Am. geophys. Un.* **64**, 90 (1983).
20. Leg 115 Shipboard Scientific Party *Nature* **329**, 586-587 (1987).
21. McDougall, I. *Geochim. cosmochim. Acta* **35**, 261-270 (1971).

## Deccan flood basalts and the Cretaceous/Tertiary boundary

V. Courtillot\*, G. Féraud†, H. Maluski‡, D. Vandamme\*, M. G. Moreau\* & J. Besse\*

\* Institut de Physique du Globe, Géomagnétisme et Paléomagnétisme (UA CNRS 729), Laboratoire de Paléomagnétisme et Géodynamique, 4 Place Jussieu, 75252 Paris Cedex 05, France

† Université de Nice, Laboratoire de Géologie et Géochimie (UA CNRS 725), Parc Valrose, 06034 Nice Cedex, France

‡ Université de Montpellier, Département de Géologie Structurale, Place Bataillon, 34060 Montpellier Cedex, France

Courtillot *et al.*<sup>1</sup> have presented palaeomagnetic, palaeontological and K-Ar data for the Deccan flood basalts which suggest that  $>10^6\text{ km}^3$  of basalt may have been erupted in  $<1$  Myr, mostly in a reversed magnetic chron. This chron is argued to be 29R, the one which contains the Cretaceous/Tertiary boundary. Here we aim to test the hypothesis<sup>1-5</sup> that the Deccan basalts could be responsible for events observed at the Cretaceous/Tertiary boundary by studying the  $^{40}\text{Ar}/^{39}\text{Ar}$  ages of samples with a wide geographic distribution. Many samples have been altered too extensively to yield plateau ages, but we have made five successful determinations. Considered with earlier<sup>6</sup> and concurrent<sup>7</sup> results, our data confirm that the bulk of the Deccan eruptions occurred in a short time, between 65 and 69 Myr, probably coincident with the Cretaceous/Tertiary boundary.

Courtillot *et al.*<sup>1</sup>, and more recently Vandamme *et al.* (manuscript in preparation), have reviewed almost one hundred K-Ar age determinations from the Deccan traps (Fig. 3). Although there is a 50 Myr spread in these ages, ages younger than



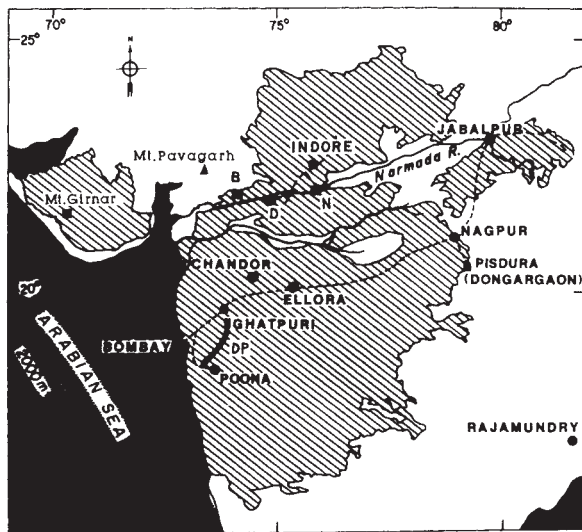


Fig. 1 Outline of the Deccan traps and location of sites sampled for  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  age determination (square symbols): B, Banganga (86 NA03); D, Dyke (86 NA18); N, Narmada, south of Indore (86 NA16, 17); Dongargaon (DK 01-03A) are from this study. Two samples from Mt Girnar and Ighatpuri were measured by Kaneoka<sup>6</sup>, and nine in the Western Ghats by Duncan and Pyle<sup>7</sup>, marked DP, between Ighatpuri and Poonma. Circles denote localities used for location.

~55 Myr have been attributed to argon loss resulting from alteration. The  $^{40}\text{Ar}$ / $^{39}\text{Ar}$  step-heating technique<sup>8</sup> can help alleviate this problem. Early determinations by Kaneoka<sup>6</sup> have been discussed by Courtillot *et al.*<sup>1</sup> and by Baksi<sup>9</sup>. Baksi concluded that only two plateau ages, for which he recalculated the  $1\sigma$  uncertainties, are acceptable: one comes from a sample near Girnar hill (Gujrat) at  $68.6 \pm 2.4$  Myr, the other from the top of the Ighatpuri section (in the western Ghats) at  $63.2 \pm 2.3$  Myr, the locations of which are shown in Fig. 1. Two plateau ages for samples near Bombay<sup>10</sup> were dismissed<sup>9</sup> because of evidence for radiogenic argon loss and overestimate of uncertainties.

The work of Courtillot *et al.*<sup>1</sup> and the dearth of critical Ar ages prompted us to undertake a new series of determinations. Samples were collected in the course of three field trips between the autumns of 1984 and 1986 on a traverse from Nagpur to Bombay<sup>1</sup> around Jabalpur and in the Narmada valley (Fig. 1).  $^{40}\text{Ar}$ / $^{39}\text{Ar}$  measurements were performed in Nice (G.F., samples 86NA03, 16, 17 and 18) and Montpellier (H.M., sample DK01-03A); the experimental procedures have been described elsewhere<sup>11,12</sup> (see Table 1 legend). The results are shown in Table 1 and Fig. 2 (ages are given at the  $2\sigma$ , or 95% uncertainty level). In all but one case, we report results for a pure plagioclase fraction extracted from the basalt to avoid the possibility of disturbed isotope ratios (which can occur in whole rock samples as a result of the different behaviour of mineralogical phases and argon retentivity). The criteria that we have used to define a plateau age are (1) the plateau region of the age spectrum should include at least 68% of the  $^{39}\text{Ar}$  released; (2) there should be at least three steps in the plateau; and (3) the individual fraction ages should agree within  $2\sigma$  with the 'integrated' age of the plateau segment. Step heating allows us to evaluate the significance of age spectra and the fairly stringent criteria outlined above ensure that the age is significant, that is, a true crystallization age. Plateau ages obtained in this way appear more reliable than isochron ages, for which criteria of data-point alignment are not so easily interpreted thermodynamically. We have relaxed rule (3) in the case of sample NA17, where one out of ten fractions differs from the plateau age by more than  $2\sigma$ . Ten whole rocks and plagioclases failed to yield acceptable plateaus. Some samples gave saddle-shaped age spectra, prob-

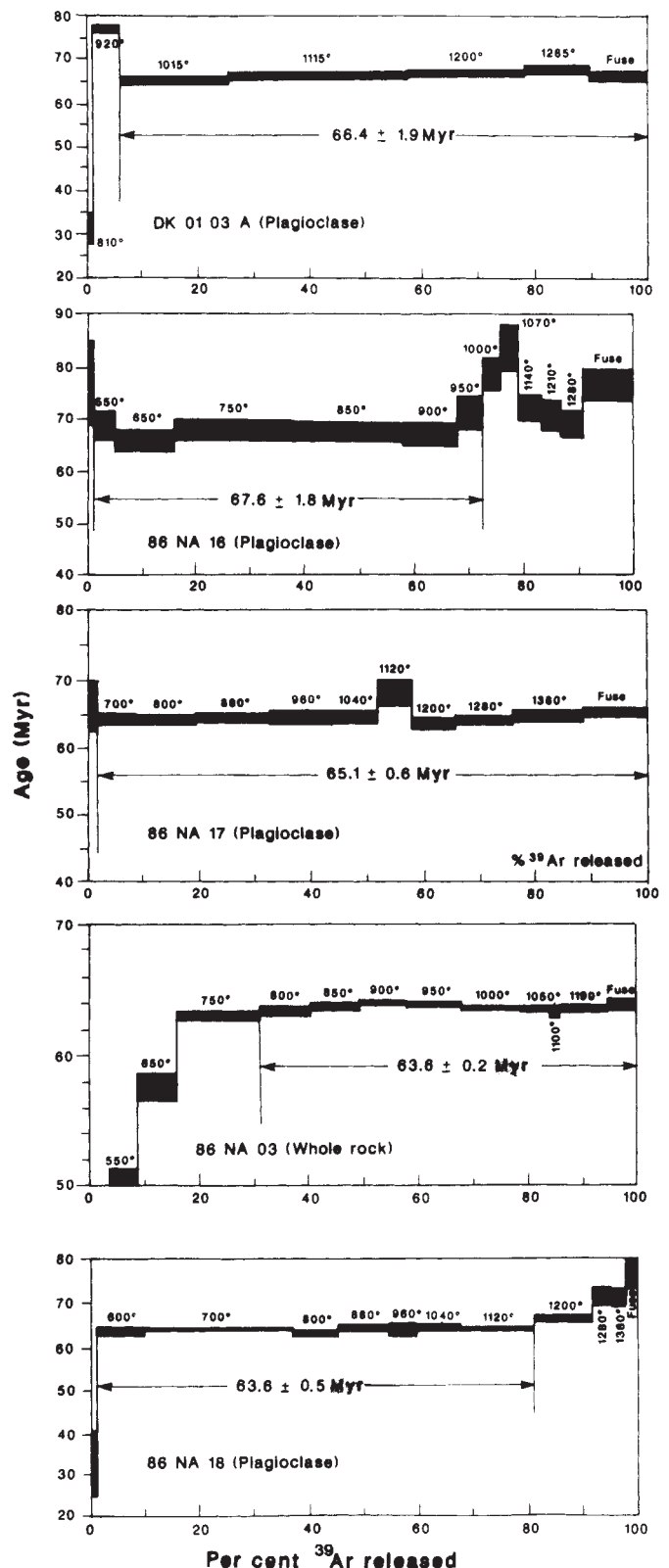
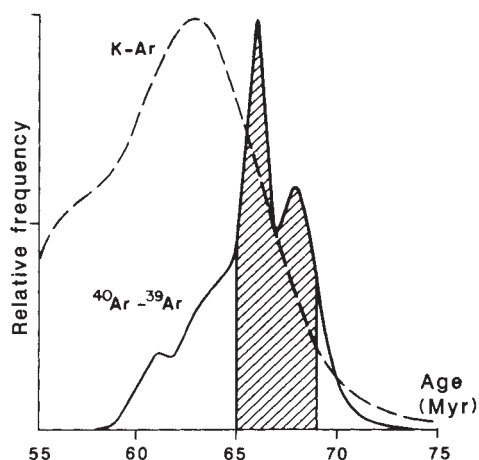


Fig. 2 Age spectra for the samples reported in the text. Uncertainties ( $1\sigma$ ) are shown for individual temperature steps. Plateau ages are given at the  $2\sigma$  (95%) uncertainty level.

ably due to excess argon. This is perhaps not surprising in view of the pervasive alteration of the Deccan traps (see ref. 13 for example) and may explain why so few determinations have so far been available.

Sample DK01-03A comes from the lowermost flow in Dongargaon (Fig. 1) and directly overlies the Maestrichtian (uppermost



**Fig. 3** Distribution of 16 reported  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  plateau ages for the Deccan traps (our data and refs 6 and 7). The histogram is shaded where relative frequency exceeds 0.4. The dashed curve is part of the corresponding distribution of  $\sim 100$  published K-Ar ages (ref. 1 and Vandamme *et al.*, manuscript in preparation).

Cretaceous) sediments of the Lameta formation<sup>14-16</sup>. Argon released between 1,015 °C and 1,500 °C leads to a consistent plateau age of  $66.4 \pm 1.9$  Myr (Fig. 2). This age corresponds to 93% of the released  $^{39}\text{Ar}$ , with only the first two temperature steps being disregarded. Samples NA16 and NA17 are from the two lowermost flows of the section in the E. Narmada valley, south of Indore (N, Fig. 1); they overlie the Bagh sediments of upper Cretaceous age<sup>17</sup>. The plagioclases NA17 and NA16 yield plateau ages of  $65.1 \pm 0.6$  Myr (98.3% of  $^{39}\text{Ar}$  released) and  $67.6 \pm 1.8$  Myr (71% of  $^{39}\text{Ar}$  released) respectively. In contrast, a dyke  $\sim 30$  m wide cutting across the traps south of the Narmada on the Indore-Bombay road (Fig. 1, site D, sample NA 18) yields a plateau age of  $63.6 \pm 0.5$  Myr, corresponding to 79.8%  $^{39}\text{Ar}$  released. Finally, we have analysed a sequence of nine plagioclase and whole rock samples from lava flows and dykes from the Banganga (Fig. 1, site B), a small tributary of the Narmada to the west of the previous sites. All but one flow (NA 03) yielded disturbed or saddle-shaped age spectra; one of the lowermost flows yielded an acceptable plateau at  $63.6 \pm 0.2$  Myr (68.5%  $^{39}\text{Ar}$  released).

During the preparation of our report<sup>18</sup>, we learned of the results obtained independently by Duncan and Pyle<sup>17</sup> from a 2,000-m composite section in the western Ghats (Ighatpury, Mahabaleshwar, Fig. 1, section DP) and sampled earlier by Beane *et al.*<sup>19</sup>. Duncan and Pyle show that there is no resolvable age difference from the bottom to the top of the series. Altogether (refs 6, 7 and this study) there are now 16 available Ar plateau age determinations for the traps. Their distribution is shown in Fig. 3; there is a very pronounced cluster of ages between 65 and 69 Myr, and a tail with a secondary peak between 60 and 65 Myr. The significance of this secondary peak (four determinations) is unclear. The Ighatpury age of Kaneoka<sup>6</sup> and the 62 Myr Neral formation age of Duncan and Pyle<sup>7</sup> can obviously be discarded for purely stratigraphic reasons, although there is no analytical indication of alteration. The 63.6 Myr age that we find in the Banganga-Narmada section (NA03) comes from a tectonically disturbed area, with unusual dips in excess of 10°, and is only 2 km from the Ambadunger intrusive carbonatite complex<sup>20</sup>. There is evidence for complete remagnetization of this flow (Vandamme *et al.*, manuscript in preparation) and for Ar loss due to alteration or a thermal event (Fig. 2). The Banganga age may therefore correspond to a younger event, limited to the north-west of the Deccan, related to opening of the Cambay graben, which can be considered as a failed rift.

**Table 1**  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  total fusion and plateau ages for the Deccan basalts

| Sample number          | Total fusion age (Myr) | Plateau age (Myr)                | $^{39}\text{Ar}$ released (%) |
|------------------------|------------------------|----------------------------------|-------------------------------|
| DK01-03A (plagioclase) | $66.5 \pm 1.9$         | $66.4 \pm 1.9$                   | 93                            |
| 86NA03 (whole rock)    | $61.0 \pm 0.3$         | $63.6 \pm 0.2$                   | 68.5                          |
| 86NA18 (plagioclase)   | $64.1 \pm 0.5$         | $63.6 \pm 0.5$                   | 79.8                          |
| 86NA16 (plagioclase)   | $69.8 \pm 1.5$         | $67.6 \pm 1.8$                   | 71.0                          |
| 86NA17 (plagioclase)   | $65.1 \pm 0.6$         | $65.1 \pm 0.6$ (step 8 excluded) | 98.3                          |

Samples were irradiated in a rotating cannister of the Mélusine reactor (CEN Grenoble), in a total flux of  $1.4 \times 10^{13} \text{ n cm}^{-2} \text{ s}^{-1}$ . The monitor used in Nice was the biotite LP6 with an age of 128.5 Myr. Samples were positioned at two or three levels in each can and three monitors were irradiated at each level. The variation in the  $^{40}\text{Ar}/^{39}\text{Ar}$  ratios for monitors at a given level was  $<0.7\%$ . The three plagioclase samples analysed in Nice were in the same can. The monitor used in Montpellier was the Caplongue Hornblende (internal standard), the age of which is  $344.5 \pm 4$  Myr. The two laboratories were interchecked and correlated. For a description of age determination and evaluation, see text and Fig. 2. Uncertainties are at the  $2\sigma$  (95%) level.

The dyke south-west of Indore (NA18) could then be a feeder of this late event, possibly associated with the opening of the Arabian Sea (ref. 1; the first oceanic anomalies observed there are 27–28, with ages of 61–64 Myr (ref. 21) or 63–65 Myr (ref. 22)).

The main  $^{40}\text{Ar}/^{39}\text{Ar}$  age histogram peak between 65 and 69 Myr is based on data with widespread distribution over the Deccan (Fig. 1). They also cover most of the stratigraphic thickness<sup>7</sup>: previous geochemical<sup>19,23</sup> and magnetostratigraphic<sup>1</sup> long-range correlations suggest that the ages found by Duncan and Pyle<sup>7</sup> should apply to the bulk of the western Ghats volcanics, namely, to the thickest part of the traps. The  $^{40}\text{Ar}/^{39}\text{Ar}$  results strongly support the contention that the bulk of Deccan volcanism occurred over  $<4$  Myr, possibly 2 Myr. Magnetostatigraphy<sup>1</sup> remains the strongest indication that the total duration may actually have been  $<1$  Myr. Also, the absolute age includes that often quoted for the Cretaceous/Tertiary boundary at 66.4 Myr (ref. 22). The case for a Cretaceous/Tertiary boundary-Deccan trap connection can certainly not be discounted. The Deccan eruption could have been the largest volcanic event in Mesozoic and Cenozoic times<sup>1,5</sup> and further study of its geochemical and ecological consequences is warranted<sup>5,24,25</sup>.

Samples were collected in autumn of 1984, 1985 and 1986 by teams including R. Chaubey, L. Gwalani, A. Prasad, R. S. Rana, A. Sahni, A. Tripathi from the universities of Punjab and Nagpur, and by H. C. Cappetta, E. Humler, J. J. Jaeger, A. Prinzhofer, J. Sudre and M. Vianey-Liaud from the universities of Paris, Montpellier and Strasbourg. We are particularly grateful to A. Sahni and L. Gwalani for help in the field, and to Mr K. P. Singh, General Manager of Gujarat Mineral Development Corporation Ltd, for kind permission to sample in the Ambadunger area.

Received 28 March; accepted 17 May 1988.

1. Courtillot, V. *et al.* *Earth Planet. sci. Lett.* **80**, 361–374 (1986).
2. Vogt, P. R. *Nature* **240**, 338 (1972).
3. McLean, D. M. in *Evolutionary Biology* Vol. 19 (ed. Hecht, M. K., Wallace, B. & Prance, G. T.) 287–313 (Plenum, New York, 1985).
4. Officer, C. B. & Drake, C. L. *Science* **219**, 1383–1390 (1983).
5. Courtillot, V. & Cisowski, S. *Eos* **68**, 193–200 (1987).
6. Kaneoka, I. *Earth planet. Sci. Lett.* **46**, 233–243 (1980).
7. Duncan, R. A. & Pyle, D. G. *Nature* **333**, 841–843 (1988).
8. Merrihue, C. & Turner, G. *J. geophys. Res.* **71**, 2852–2857 (1966).
9. Baksi, A. K. *Geology* **15**, 147–150 (1987).
10. Venkatesan, T. R., Pande, K. & Gopalan, K. *J. geol. Soc. India* **27**, 102–109 (1986).
11. Féraud, G., Gastaud, J., Auzende, J. M., Olivet, J. L. & Cornen, G. *Earth planet. Sci. Lett.* **57**, 211–226 (1982).



12. Costa, S. & Maluski, H. *Chem. Geol. isotop. Sci.* **72**, (in the press).
13. Mahoney, J. J. in *Continental Flood Basalts* (ed. MacDougall, J. D.) (Reidel, Dordrecht, in the press).
14. Buffetaut, E. *Newsl. Stratigr.* **18**, 1-6 (1987).
15. Prasad, G. V. R., Sahni, A. & Gupta, J. *Geosci. J.* **7**, 163-180 (1986).
16. Vianey-Liaud, M., Jain, S. L. & Sahni, A. *J. vert. Paleont.* **7**, 408-424 (1987).
17. Sahni, A. *Mém. Soc. Géol. France* **147**, 125-137 (1984).
18. Féraud, G. *et al. Eos* **69**, 341 (1988).
19. Beane, J. E., Turner, C. A., Hooper, P. R., Subbarao, K. V. & Walsh, J. N. *Bull. volcan.* **48**, 1-33 (1986).
20. Gwalani, L. G. *Geol. Surv. Ind. Spec. Pub.* **12**, 319-325 (1984).
21. Cox, A. V. *A Geologic Time Scale* (ed. Harland, W. B. *et al.*) 63-84 (Cambridge University Press, 1982).
22. Berggren, W. A., Kent, D. V., Flynn, J. F. & Van Couvering, J. A. *Geol. Soc. Am. Bull.* **96**, 1407-1418 (1985).
23. Devey, C. W. & Lightfoot, P. C. *Bull. volcan.* **48**, 61-83 (1986).
24. Courtillot, V. *Eos* **69**, 301 (1988).
25. Courtillot, V. & Besse, J. *Science* **237**, 1140-1147 (1987).

## Post-feeding thermotaxis and daily vertical migration in a larval fish

Wayne A. Wurtsbaugh & Darcy Neverman\*

Department of Fisheries and Wildlife/Ecology Center,  
Utah State University, Logan, Utah 84322-5210, USA

Many aquatic animals make daily vertical migrations, typically ascending into warm shallow strata for the night and descending to cooler, deeper layers of lakes or oceans for the day. Although some organisms may migrate to avoid predation<sup>1-3</sup> researchers have also suggested that daily migration is a thermoregulatory strategy allowing ectotherms to lower their metabolic rates in cold, deep waters, thus conserving energy<sup>4,5</sup>. Tests of this hypothesis, however, have been equivocal<sup>6-8</sup>. Here we suggest an alternative hypothesis: that fish ascend into warmer water after feeding to stimulate digestion, thereby allowing greater feeding and growth. We tested this hypothesis using the Bear Lake sculpin (*Cottus extensus*) which feeds on the bed of the lake during the day, and at night migrates into the water column where temperatures are 10 °C warmer. The warmer temperatures promoted digestion and allowed the fish to feed and grow three times faster than if they had remained in the cold hypolimnion. Thus, daily vertical migration in this species is an adaptation allowing them to exploit thermal gradients in their environment<sup>9</sup> to maximize energetic intake.

Our field study was conducted in Bear Lake, Utah-Idaho, USA (42°00' N, 111°20' W), a 520 km<sup>2</sup>, oligotrophic system, with a maximum depth of 60 m. The lake is thermally stratified in the summer, with surface temperatures near 20 °C and bottom temperatures of 4-5 °C. We studied the movements of juvenile (6-30 mm standard length) sculpin during their first summer of life.

The juvenile fish had a pronounced daily vertical migration. During the daylight hours they were on the bottom where temperatures were 4-5 °C (Fig. 1a). At night, however, they moved into the water column, appearing at dusk and disappearing at dawn (Fig. 1b). When they moved into the water column the fish concentrated in the region of the thermocline (metalimnion) at temperatures between 13-16 °C (Fig. 2).

Although many organisms migrate into surface waters to feed<sup>1-3</sup>, the sculpin did not consume prey when they were in the water column. Analysis of 330 sculpin collected throughout the summer demonstrated that they consumed primarily ostracods (59% by weight) and adult cyclopoid copepods (39%). These organisms were found only in the sediments or immediately above them (D. Neverman, unpublished data). Furthermore, the sculpin fed only during the day when they were on the bottom, and stomach filling peaked just before the

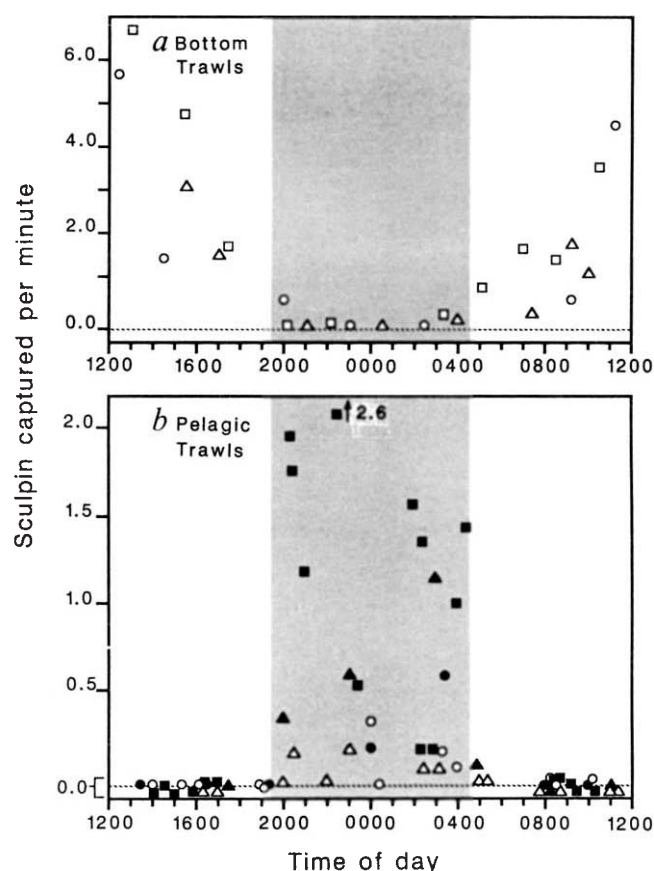


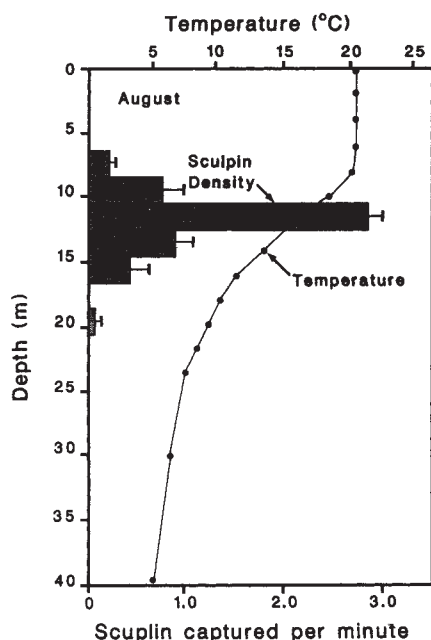
Fig. 1 Temporal changes in the catch rates of juvenile Bear Lake sculpin collected with trawls on the bottom of the lake (a), and in the water column (b). Closed symbols in b indicate captures in the metalimnion; open symbols indicate densities in the epilimnion and hypolimnion. Circles, 30-31 July 1985; triangles, 14-15 August 1985; squares, 27-28 August 1985. The shaded area denotes the time from evening to morning civil twilight (Mountain standard time) on 14 August. Fish were sampled at a site 5.5 km from shore where the mean depth was 40 m. We collected fish on the bottom with a 2.9-m otter trawl pulled at 1.0 m s<sup>-1</sup>. In the water column they were sampled with a 1-m<sup>2</sup> opening-closing Tucker trawl pulled at 1.5 m s<sup>-1</sup>. Although capture efficiencies during the day may have been lower than at night due to visual net avoidance, catch rates of adult sculpin in the otter trawl were not significantly different between day and night (*t*-test; *P* = 0.338). This indicates that changes in avoidance of this net were minimal during any 24-h period.

dusk ascent (Fig. 3). During the night the fullness of the gut decreased monotonically, reaching a minimum at dawn as the fish returned to the bottom.

We reasoned that the fish moved to warmer strata to increase their digestion rate. We tested this in the laboratory by measuring gut clearance rates of satiated sculpin at two temperatures. At 5 °C the instantaneous digestion rate<sup>10</sup> was 3.2% per hour ( $\pm 1.1$ ; 95% confidence interval, CI), and only 22% of the meal was evacuated from the stomach during the 7.5 h digestion trial. At this rate, 50 hours would have been required for a fish to digest 80% of its meal. In contrast, at 15 °C the digestion rate was 23.4% per hour ( $\pm 0.9$ ; 95% CI) and 80% of the meal was evacuated in 7.5 hours. This rate was nearly identical to actual digestion rates of wild fish (25-27% per hour) captured in the metalimnion (Fig. 3). Nocturnal migration into warmer water apparently allowed large meals to be digested rapidly before feeding began the following day.

The rapid evacuation of the gut during the night also increased the feeding rate the following day. Sculpin that had digested their meals at 15 °C and were then fed at 5 °C made significantly

\* Present address: Missouri Department of Conservation, 2500 S. Halliburton, Kirksville, Missouri 63501, USA.



**Fig. 2** Nighttime distribution of juvenile Bear Lake sculpin in the water column in relation to temperature and depth. Error bars indicate  $\pm 1$  s.e. of catch rates of 2–3 replicate Tucker trawls. Collections were made between 2200 and 0400 hours (Mountain standard time) on 8–9 August 1986. Water temperatures were measured with a thermistor.

more feeding attempts than fish kept continuously at 5 °C (Table 1;  $P < 0.026$ , paired  $t$ -test).

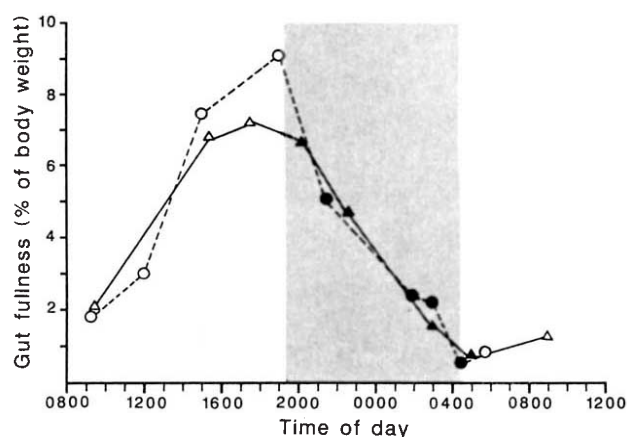
This increased feeding rate would be advantageous only if it provided sufficient energy to overcome the elevated respiratory costs of inhabiting warm water at night. To determine if a migration could increase the net energy gain of sculpin, we

**Table 1** Effects of constant and fluctuating temperatures on feeding and growth rates (dry weight) of juvenile Bear Lake sculpin (*Cottus extensus*) reared in the laboratory

| Temperature | Condition simulated | Feeding rate* at 5 °C (strikes per fish min <sup>-1</sup> ) | Growth rate* (% per day) |
|-------------|---------------------|-------------------------------------------------------------|--------------------------|
| 5 °C        | Hypolimnion         | 0.22 $\pm$ 0.08                                             | 0.74 $\pm$ 0.30          |
| 5–15 °C     | Migratory           | 0.48 $\pm$ 0.15                                             | 2.15 $\pm$ 0.45          |

To quantify feeding rates in constant and fluctuating temperatures we first allowed groups of ten fish to digest a large meal of zooplankton overnight (7.5 h) at either 5 or 15 °C. Fish at 15 °C were then moved to 5 °C and allowed 1 h to acclimate. We then fed the sculpin live zooplankton and observed the number of feeding strikes made at prey during six 2-min periods over the following hour. To measure growth, fish with an estimated mean dry mass of 2.4 mg were anesthetized, measured, and randomly assigned to a temperature treatment (5 °C; or 5–15 °C, 16 h:8 h), or killed to provide a subsample for determining the initial length-dry-weight relationship. In each treatment we reared 15 sculpin individually in 0.5-litre containers constructed of 300 and 1,000- $\mu$ m nylon mesh on the sides and bottom, respectively. We placed these in glass, flow-through aquaria. To simulate natural conditions, fish were fed repletion rations of plankton only during the day. The diet was 75% *Ceriodaphnia* sp., 8% *Daphnia magna*, and 7% *Cylops* spp., all between 300–500  $\mu$ m. During the day all sculpin were kept at 5 °C. In the evening the cups were lifted from their tanks, rinsed to remove plankton, and placed into a zooplankton-free aquaria of the appropriate night time temperature. After eight hours we moved the fish again and allowed them one hour in their new aquaria before feeding. The temperature of the 5–15 °C treatment was tempered over 1-h periods when the fish were transferred. After 21 days, the fish were removed, dried and weighed on a microbalance. Other experimental protocol and growth calculations followed Wurtsbaugh and Cech<sup>17</sup>.

\* Means  $\pm$  95% confidence intervals.



**Fig. 3** Temporal changes in the fullness of the guts of juvenile sculpin captured in Bear Lake. Closed symbols are from fish collected in the water column at night. Open symbols are from fish captured on the bottom of the lake. Each point represents the mean for ten sculpin. The shaded area indicates the time from evening to morning civil twilight on 14 August 1985. Triangles: 14–15 August 1985. Circles: 16–17 September 1985. Instantaneous digestion rates<sup>10</sup> of sculpin during the night were calculated from the exponential decline in the weight ( $W$ ) of food in the guts over time ( $t$ ): August,  $W_{t2} = W_{t1} e^{-0.27(t)}$ ,  $r^2 = 0.92$ ; September,  $W_{t2} = W_{t1} e^{-0.25(t)}$ ,  $r^2 = 0.92$ .

conducted a growth experiment in the laboratory. Fish were exposed to either a constant temperature of 5 °C (bottom temperatures), or a fluctuating regime of 5–15 °C that mimicked the actual temperatures encountered by migrating individuals. To simulate the actual feeding regime, fish were fed *ad libitum* only during the day.

Sculpin in the fluctuating temperature regime grew 300% faster than those reared at 5 °C. This difference was highly significant ( $P < 0.0001$ ; Table 1). The growth experiment, however, may have over-estimated the net energy benefit of the migratory behaviour of wild sculpin, because fish in the laboratory did not have to swim to the metalimnion and maintain themselves in the water column each night. Swimming costs of small fish like sculpin have not been measured, but moderate swimming speeds of 1–2 body lengths per second decrease growth rates less than 25% in well-fed juvenile salmon<sup>11</sup>. Consequently, the metabolic costs of routine swimming are probably not so large as to cancel the threefold growth advantage of inhabiting the warm metalimnion at night.

Post-feeding thermotaxis may be a common phenomenon affecting not only sculpin, but a wide range of ectothermic animals. Under experimental conditions some reptiles, amphibians and other species of fish seek higher temperatures after feeding<sup>12,13</sup>. Also, bioenergetic studies on reptiles<sup>14,15</sup> and amphibians<sup>16</sup> have shown that post-feeding thermotaxis increases their feeding, digestion and growth. Our results indicate that daily vertical migration and behavioural thermoregulation can also be important in controlling growth processes of fish.

We thank S. Shope for his assistance in the field, C. Hawkins, D. Levy and J. MacMahon for reviewing the manuscript, and N. Banks for careful editing. The Utah Division of Wildlife Resources provided logistical support and the Utah Cooperative Fisheries and Wildlife Research Unit contributed personal assistance. Funded by the Ecology Center and the Vice-President for Research of Utah State University.

Received 15 April; accepted 17 May 1988.

- Gliwicz, J. Z. *Nature* **320**, 746–748 (1986).
- Sijch, H. B. & Lampert, W. *Nature* **293**, 396–398 (1981).
- Clark, C. W. & Levy, D. A. *Am. Nat.* **131**, 271–290 (1988).
- McLaren, I. A. *J. Fish. Res. Bd Can.* **20**, 685–727 (1963).
- Brett, J. R. *Am. Zool.* **11**, 99–113 (1971).



6. Lock, A. R. & McLaren, I. A. *Limnol. Oceanogr.* **15**, 638-640 (1970).
7. Swift, M. C. *Ecology* **57**, 900-914 (1976).
8. Biette, R. M. & Geen, G. H. *Can. J. Fish. Aquatic Sci.* **37**, 203-210 (1980).
9. Crowder, L. B. & Magnusson, J. J. in *Behavioral Energetics: The Cost of Survival in Vertebrates* (eds Aspey, W. P. & Lustick, S. L.) 189-221 (Ohio State Univ. Press, Columbus, 1983).
10. Persson, L. in *Contemporary Studies on Fish Feeding* (eds Simenstad, C. A. & Cailliet, G. M.) 51-58 (Junk, Dordrecht, 1986).
11. White, J. R. & Li, H. W. *Comp. Biochem. Physiol.* **81A**, 25-33 (1985).
12. Huey, R. B. in *Biology of the Reptilia* Vol. 12 (eds Gans, C. & Pough, F. H.) 25-91 (Academic, New York, 1982).
13. Reynolds, W. W. & Casterlin, M. E. *Am. Zool.* **19**, 211-224 (1979).
14. Avery, R. A. *Symp. zool. Soc. Lond.* **52**, 407-424 (1984).
15. Christian, K. A. *Can. J. Zool.* **64**, 836-840 (1986).
16. Lillywhite, H. B., Licht, P. & Chelgren, P. *Ecology* **54**, 375-383 (1973).
17. Wurtsbaugh, W. A. & Cech, J. J. *Trans. Am. Fish. Soc.* **112**, 653-660 (1983).

## Chronic ethanol causes heterologous desensitization of receptors by reducing $\alpha_s$ messenger RNA

Daria Mochly-Rosen\*§, Fu-Hsiung Chang†, Laura Cheever§, Mark Kim§, Ivan Diamond\*†‡§ & Adrienne S. Gordon\*†§

Departments of Neurology\*, Pharmacology† and Pediatrics‡, and the Ernest Gallo Clinic and Research Center§, University of California School of Medicine at San Francisco, San Francisco General Hospital, San Francisco, California 94110, USA

One of the biochemical results of ethanol exposure is a change in the amount of the intracellular second messenger cyclic AMP (cAMP) produced in response to receptor stimulation<sup>1-11</sup>. In general, acute ethanol exposure increases the amount of cAMP produced on stimulation of receptors coupled to the enzyme adenylyl cyclase via the GTP-binding protein  $G_s$ <sup>1,4-8,10</sup>, whereas chronic ethanol exposure has the opposite effect<sup>1,2,7,9,10</sup> (results for receptors coupled via  $G_i$  have been more variable). We previously reported that adaptation to continuous ethanol exposure reduces receptor-stimulated cAMP production by 25-35% in a neuroblastoma cell line (NG108-15)<sup>1</sup>, and an even greater reduction of 75% was observed in lymphocytes taken from actively-drinking alcoholics<sup>2</sup>. This reduction in receptor-stimulated cAMP levels was recently confirmed in platelets from alcoholics<sup>12</sup>. None of these studies, however, determined whether more than one receptor coupled to adenylyl cyclase activity was affected in the same cell. Here we report that chronic ethanol exposure causes desensitization of heterologous receptors coupled to  $G_s$  as cAMP production mediated by prostaglandin E<sub>1</sub> as well as by adenosine is reduced by ~30% in NG108-15 cells. We show that, after chronic ethanol exposure, the activity of the  $\alpha$  subunit of  $G_s$  is decreased by 29%, the amount of  $\alpha_s$  protein is decreased by 38.5%, and  $\alpha_s$  messenger RNA is decreased by 30%. Thus, cellular adaptation to ethanol involves a reduction in  $\alpha_s$  mRNA and, as a consequence, reduced cAMP production by heterologous receptors coupled to  $G_s$ . Such changes in cAMP production may account for the tolerance and physical dependence on ethanol in alcoholism.

Sharma *et al.*<sup>13</sup> were the first to use neural cells in culture as a model system to study the role of adenylyl cyclase in the development of tolerance and dependence to opiates. We and others<sup>5,7,9</sup> have used cells in culture as model systems to examine cellular adaptation to ethanol under well-defined conditions. In NG108-15 neuroblastoma × glioma cells, treatment with 100 mM ethanol for 48 h reduces adenosine receptor-stimulated cAMP levels by 25-35% when assayed in the absence of ethanol (ref. 1, Table 1). This is consistent with the data on cAMP production in lymphocytes<sup>2</sup> and platelets<sup>12</sup> from actively drinking alcoholics. Using the NG108-15 cell line, which has both adenosine and prostaglandin E<sub>1</sub> (PGE<sub>1</sub>) receptors that activate adenylyl cyclase, we could determine whether ethanol affects more than one receptor in the same cell. We found that PGE<sub>1</sub>

receptor-stimulated cAMP production was decreased by 25.6% in this cell line after 48 h exposure to 100 mM ethanol, similar to the decrease in adenosine receptor-mediated cAMP production (Table 1). This result suggests that chronic ethanol treatment causes heterologous desensitization of receptors that stimulate adenylyl cyclase activity.

Production of cAMP requires at least three membrane components<sup>14,15</sup>: receptors for hormones and neurotransmitters, a GTP-binding protein,  $G_s$ , and the enzyme that synthesizes cAMP, adenylyl cyclase<sup>14</sup>. Alterations in  $G_s$  could account for the reduced stimulation of cAMP production by both adenosine and PGE<sub>1</sub>. Therefore, we investigated whether chronic ethanol treatment alters the amount and/or function of  $\alpha_s$ , the subunit of  $G_s$  that mediates receptor stimulation of adenylyl cyclase.

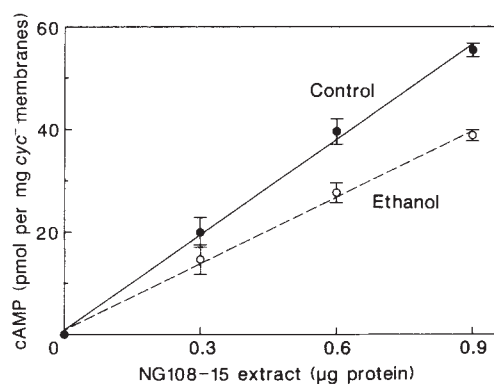
To assay  $\alpha_s$  function, S49 lymphoma *cyc*<sup>-</sup> cells were reconstituted with the  $\alpha_s$  subunit in a Lubrol extract of membranes from control and from ethanol-treated cells. Cells that are *cyc*<sup>-</sup> lack  $\alpha_s$  and therefore cannot activate adenylyl cyclase via  $\beta$ -adrenergic receptor stimulation<sup>16</sup> unless supplemented with exogenous  $\alpha_s$ . The amount of cAMP produced in this system in the presence of the agonist isoprenaline and a non-hydrolysable GTP analogue, GppNHp thus reflects the amount and/or the ability of the added  $\alpha_s$  to couple receptor activation to adenylyl cyclase. Consistent with the decrease in adenosine- and PGE<sub>1</sub>-mediated cAMP production in intact cells, chronic ethanol treatment caused a 29% reduction in the ability of membrane extracts to complement the  $\alpha_s$  deficiency of *cyc*<sup>-</sup> membranes (Fig. 1 and Table 1). This result was not due to any alteration in  $\alpha_s$  extraction efficiency after chronic ethanol treatment (see legend to Fig. 1).

Decreased cAMP production in the reconstitution assay with *cyc*<sup>-</sup> membranes could result from a reduced coupling of the  $\alpha_s$  to either adenylyl cyclase, or the  $\beta$ -adrenergic receptor or both. Alternatively, decreased amounts of  $\alpha_s$  protein in the ethanol-treated cells could be responsible. Using anti- $\alpha_s$  antibodies in a Western blot analysis, we measured membrane-bound  $\alpha_s$  in control and ethanol-treated cells. Figure 2 shows that chronic exposure to ethanol results in a decrease in  $\alpha_s$  protein; densitometric analysis of the autoradiograms revealed the decrease was 38.5% (Table 1). The fact that the reductions

**Table 1** The effect of chronic ethanol exposure on  $\alpha_s$  activity and synthesis

|                                           | Decrease (%)       |
|-------------------------------------------|--------------------|
| cAMP production                           |                    |
| Adenosine receptor-dependent              | 27.0 ± 6.0 (n = 4) |
| PGE <sub>1</sub> receptor-dependent       | 25.6 ± 5.7 (n = 3) |
| $\alpha_s$ Function                       | 29.0 ± 1.0 (n = 3) |
| ( <i>cyc</i> <sup>-</sup> reconstitution) |                    |
| $\alpha_s$ Protein                        | 38.5 ± 5.5 (n = 4) |
| (Western blot analysis)                   |                    |
| $\alpha_s$ mRNA                           | 30.0 ± 2.0 (n = 4) |
| (Northern blot analysis)                  |                    |

NG108-15 cells were treated with 100 mM ethanol for 48 h<sup>1</sup>. Adenosine receptor-dependent cAMP levels were determined in intact control and ethanol-treated cells using phenylisopropyladenosine as described<sup>1</sup>. cAMP assays, Western and Northern blot analyses were carried out on the same cells. 'n' indicates the number of independent experiments on cell cultures grown at different times. The amount of  $\alpha_s$  mRNA was also determined by slot blot analysis using 0.625, 1.25, 2.5 and 5.0 µg of total RNA. The results are from four separate RNA preparations analysed on four Northern blots and one slot blot. Western, Northern and slot blots were analysed by scanning autoradiograms with a Hoefer GS 300 densitometer connected to a Hewlett Packard 3390A integrator. Each band was analysed by scanning four separate segments and the areas from each scan were added to compensate for variations in band width across the lane. Results are expressed as percent decrease (mean ± s.e.m.) of ethanol-treated cells as compared to controls.

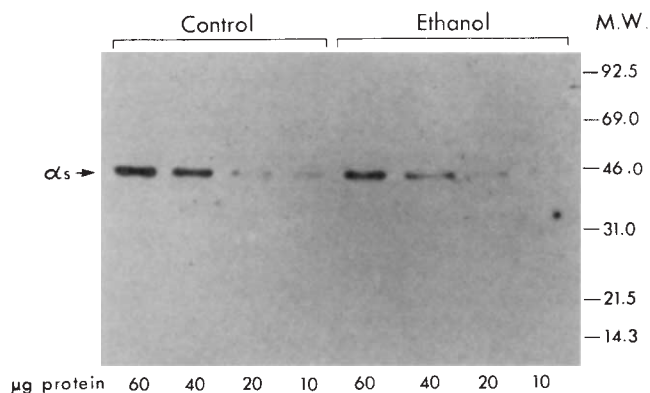


**Fig. 1** Activity of  $\alpha_s$  in NG108-15 membranes measured by reconstitution of  $\alpha_s$ -deficient  $cyc^-$  membranes. Increasing amounts of Lubrol PX extracts from NG108-15 membranes isolated from control (●) and ethanol-treated cells (○) were incubated with  $cyc^-$  membranes, and receptor-stimulated adenylyl cyclase activity was measured.

**Methods.** Ethanol-treated and control cells<sup>1</sup> were collected and washed in Tris-buffered saline (20 mM Tris-HCl, 150 mM NaCl, pH 7.4) and subjected to hypotonic shock in 5 mM Tris-HCl, 5 mM MgCl, pH 7.4 in the presence of 20  $\mu\text{g ml}^{-1}$  leupeptin and 0.1 mM dimethyl sulphoxide (DMSO) for 15 min on ice, followed by disruption by sequential passage through 23 and 26 gauge needles. Nuclei and debris were removed by centrifugation for 5 min at 500g and the supernatants were stored at  $-70^\circ\text{C}$ . Membranes were thawed and centrifuged at 170,000g for 20 min, washed once in 50 mM HEPES, pH 7.0, resuspended at 1 mg protein  $\text{ml}^{-1}$  and extracted with 0.2% Lubrol PX. Activity of  $\alpha_s$  in the extract was measured by correcting the  $\alpha_s$  deficiency of S49 lymphoma  $cyc^-$  membranes in a complementation assay. Western blot analysis using anti- $\alpha_s$  antibodies verified that Lubrol extraction was complete, with no detectable  $\alpha_s$  remaining in the extract pellet obtained by centrifugation at 170,000g for 20 min. Moreover, no  $\alpha_s$  was found in the cytosol of control and ethanol-treated cells. The  $\alpha_s$  reconstitution assay was carried out as described<sup>16</sup>. The total amount of protein was kept constant by adding heat-inactivated extract<sup>18</sup>. The reaction was stopped after 20 min by addition of 250  $\mu\text{l}$  of cold 0.1 N HCl. cAMP was measured by radioimmunoassay<sup>20</sup> and is expressed as pmol per mg  $cyc^-$  membrane protein per 20 min. Results are the mean  $\pm$  s.e.m. of three separate experiments carried out in triplicate.

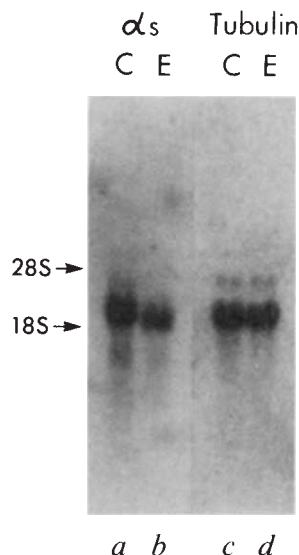
in adenosine- and  $\text{PGE}_1$ -stimulated cAMP production, in the functional reconstitution assay, and in  $\alpha_s$  protein (Table 1) were comparable, suggests that the reduction in the amount of  $\alpha_s$  protein is sufficient to account for the decrease in receptor-stimulated cAMP production. The amount of  $\alpha_i$ , the GTP-binding subunit of the inhibitory G protein, is not altered in these cells by chronic ethanol exposure (M. E. Charness, personal communication). It appears, therefore, that the absolute amount of  $\alpha_s$  limits cAMP production in NG108-15 cells. This is consistent with a recent observation in the GH<sub>3</sub> rat pituitary cells line<sup>17</sup>, where both  $\alpha_s$  protein and adenylyl cyclase activity increased to the same extent (twofold) on treatment with dexamethasone.

To investigate the mechanism underlying the ethanol-induced reduction in  $\alpha_s$  protein, we compared the amount of  $\alpha_s$ -specific mRNA in ethanol-treated and control cells. Northern blot analysis of total cellular RNA<sup>18</sup> showed a 30% reduction in  $\alpha_s$  mRNA in the ethanol-treated cells (Fig. 3a, b and Table 1). In contrast,  $\beta$ -tubulin mRNA was slightly increased in the same experiments (Fig. 3c, d). Chronic ethanol has also been shown to cause a 20–40% decrease in mRNA for proopiomelanocortin<sup>11</sup> and a twofold increase in mRNA for the major histocompatibility complex<sup>19</sup>. Thus, the ethanol-induced reduction in mRNA coding for  $\alpha_s$  is unlikely to be due to global changes in cellular mRNA. It remains to be determined whether decreased mRNA



**Fig. 2** Western blot analysis of  $\alpha_s$  protein.

**Methods.** Equal amounts of membrane protein from control and ethanol-treated cells were electrophoresed on SDS-polyacrylamide gels, electrotransferred to nitrocellulose paper, and then probed with rabbit anti- $\alpha_s$  antiserum (serum A-572, a generous gift from Drs Susanne M. Mumby and Alfred G. Gilman) as described<sup>21</sup>.



**Fig. 3** Northern blot analysis of  $\alpha_s$  and  $\beta$ -tubulin mRNA. Total RNA was isolated from control (a, c) and ethanol-treated (b, d) cells. RNA was electrophoresed on formaldehyde-agarose gels and probed with radioactive rat  $\alpha_s$  and human  $\beta$ -tubulin complementary DNA probes.

**Methods.** Total RNA (20  $\mu\text{g}$ ), extracted from control (a, c) and ethanol-treated (b, d) cells by the guanidine thiocyanate-LiCl method<sup>17,22</sup>, was electrophoresed on 1% agarose-formaldehyde<sup>17,23</sup> and transferred to Zetaprobe (Bio-Rad) for Northern analysis as described<sup>18</sup>.  $^{32}\text{P}$ -dATP-labelled cDNA probes for  $\alpha_s$  (a, b) and  $\beta$ -tubulin (c, d) were used to measure the corresponding mRNA as described<sup>17,24</sup>. A cDNA probe for rat  $\alpha_s$  (carboxy-terminal two thirds of the open reading frame) was obtained from Dr Randall Reed. cDNA probe for human  $\beta$ -tubulin (entire coding region) was obtained from Dr Larry Kedes.

coding for  $\alpha_s$  is due to a reduction in transcription and/or the stability of  $\alpha_s$  mRNA.

In summary, treatment with 100 mM ethanol for 48 h results in a 27% reduction in adenosine-stimulated cAMP levels and a 25.6% reduction in  $\text{PGE}_1$ -stimulated cAMP levels in NG108-15 cells. We have also found a comparable decrease in mRNA for  $\alpha_s$  and in the amount of  $\alpha_s$  protein in the same cells. Reduced amounts of  $\alpha_s$  protein may account for the 75% decrease in adenosine receptor-stimulated cAMP levels we find in lymphocytes from alcoholics<sup>2</sup>. Heterologous receptor desensitization consequent on a reduction in mRNA for  $\alpha_s$  could be a



critical molecular event in the pathogenesis of alcoholism and may explain many of the pleiotropic effects of ethanol on cellular functions.

We thank Dr Henry Bourne and Dr Michael Miles for helpful suggestions and discussions, Ms K. Collier and Ms M. Panigbatan for technical assistance, and Ms R. C. Webb for manuscript preparation. This work was supported in part by grants from NIAAA and the Alcoholic Beverage Medical Research Foundation.

Received 23 February; accepted 18 May 1988.

- Gordon, A. S., Collier, K. & Diamond, I. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2105-2108 (1986).
- Diamond, I., Wrubel, B., Estrin, W. & Gordon, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1413-1416 (1987).
- Volicer, L., Mirin, R. & Gold, B. I. *J. Stud. Alcohol* **38**, 11-24 (1977).
- Hynie, S., Lanefelt, F. & Fredholm, B. B. *Acta pharmac. tox.* **47**, 58-65 (1980).
- Rabin, R. A. & Molinoff, P. B. *J. Pharmac. exp. Ther.* **227**, 551-556 (1983).
- Luthin, G. R. & Tabakoff, B. *J. Pharmac. exp. Ther.* **228**, 579-587 (1984).
- Richelson, E., Stenstrom, S., Forray, C., Enloe, L. & Pfenning, M. *J. Pharmac. exp. Ther.* **239**, 687-692 (1986).
- Saito, T., Lee, J. M. & Tabakoff, B. *J. Neurochem.* **44**, 1037-1044 (1985).
- Charness, M. E., Querimitt, L. A. & Diamond, I. *J. biol. Chem.* **261**, 3164-3169 (1986).
- Bauche, F., Bourdeaux-Jaubert, A. M., Guidicelli, Y. & Nordmann, R. *FEBS Lett.* **219**, 296-300 (1987).
- Dave, J. R., Eiden, L. E., Karanian, J. W. & Eskay, R. L. *Endocrinology* **118**, 280-286 (1986).
- Tabakoff, B. *et al. N. Engl. J. Med.* **318**, 134-139 (1988).
- Sharma, S. K., Klee, W. A. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3092-3096 (1975).
- Stryer, L. & Bourne, H. R. *A. Rev. Cell Biol.* **2**, 391-419 (1986).
- Birnbaumer, L. *et al. in Molecular Mechanisms of Transmembrane Signalling* (eds Cohen, P. & Houslay, M. D.) 131-182 (Elsevier, New York, 1985).
- Farfel, Z., Brickman, A. S., Kaslow, H. R., Brothers, V. M. & Bourne, H. R. *N. Engl. J. Med.* **303**, 237-242 (1980).
- Chang, F.-H. & Bourne, H. R. *Endocrinology* **121**, 1711-1715 (1987).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Parent, L. J., Ehrlich, R., Matis, L. & Singer, D. S. *FASEB J.* **1**, 469-473 (1987).
- Koch, T. K., Gordon, A. S. & Diamond, I. *Biochem. Biophys. Res. Commun.* **114**, 339-347 (1983).
- Mumby, S. M., Kahn, R. A., Manning, D. R. & Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 265-269 (1986).
- Cathala, G. *et al. DNA* **2**, 329-335 (1983).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).

## Regulatory properties of LFA-1 $\alpha$ and $\beta$ chains in human T-lymphocyte activation

Carel van Noesel, Frank Miedema, Miranda Brouwer, Menno A. de Rie, Lucien A. Aarden & René A. W. van Lier\*

Central Laboratory of the Netherlands Red Cross Blood Transfusion Service and Laboratory for Experimental and Clinical Immunology, University of Amsterdam, Amsterdam, The Netherlands

Lymphocyte function-associated antigen-1 (LFA-1) is a heterodimer composed of an  $\alpha$  and a  $\beta$  chain that is expressed on the surface of most leukocytes and is an essential molecule for adhesion reactions between cells participating in the immune response<sup>1,2</sup>. A putative ligand for LFA-1 is the intercellular adhesion molecule ICAM-1 (refs 3-5). Leukocyte adhesion abnormality is found in patients with LFA-1 deficiency<sup>6-7</sup>. It is not clear whether binding of ligand to the LFA-1 molecule merely spatially orientates cells towards each other or can also induce signals that regulate cell activation and differentiation. We have recently developed a T-cell proliferation assay which uses immobilized anti-CD3 monoclonal antibodies as stimulant and is independent of LFA-1-mediated cellular adhesion. As there is no interference by anti-LFA-1 monoclonal antibodies with the adhesion-dependent activation steps, this T-cell activation system allows us to investigate whether transmembrane signals are induced by binding of ligand to LFA-1 on T cells. Our data indicate that binding of

ligand to LFA-1 results in the transduction of regulatory signals across the plasma membrane, rather like other molecules (CD2, CD4, CD8) (refs 8-11) with signal-modifying properties involved in the adhesion of T cells to target/stimulator cells. Indeed, adhesion molecules might generally be important in signal transduction, even in cells not belonging to the immune system.

Proliferation of human T cells can be induced by antibodies directed against the CD3 molecule. This activation model mimics antigen-specific T-cell triggering and is critically dependent on the presence of monocytes as accessory cells<sup>12</sup>. Monocytes participate in T-cell stimulation by immobilizing anti-CD3 through their Fc receptors, as well as by the production of lymphokines such as IL-1 (ref. 13). The requirement for monocytes can be circumvented by immobilization of anti-CD3 antibody on a solid phase. A T-cell activation assay has recently been described<sup>14,15</sup> in which anti-CD3 antibodies are immobilized on the polystyrene of flat-bottomed culture wells. The induced T-cell proliferation is proportional to the density of the immobilized anti-CD3 antibody and is accessory-cell (monocyte)-independent<sup>15</sup> (Fig. 1a). As LFA-1-mediated T-cell accessory-cell interaction is an essential requirement for T-cell proliferation in conventional culture systems, T cells from patients with leukocyte adhesion deficiency (LAD) proliferate poorly in response to most stimuli<sup>7</sup>. We found that soluble anti-CD3 antibody could not stimulate peripheral blood lymphocytes of an LAD patient to proliferate (Table I). Monocytes of a healthy donor could not restore proliferation, which indicates that LFA-1 expression on T cells is essential for proliferation induction by soluble anti-CD3. But, when the T cells of this patient were activated with immobilized anti-CD3, the extent of proliferation was comparable (Fig. 1b) with that induced in healthy donor T cells. There were qualitative but not quantitative differences in the proliferation of normal T cells and T cells of an LAD patient: T cells from healthy donors

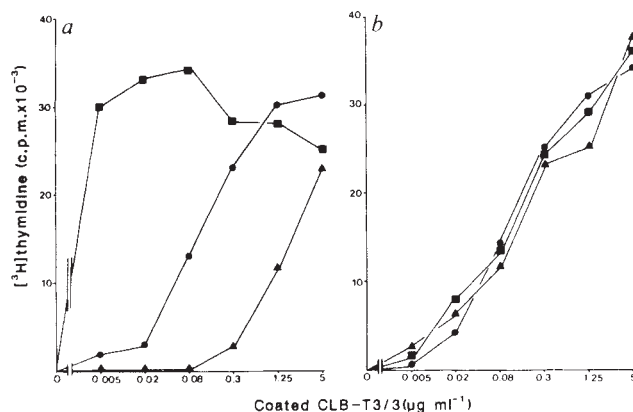


Fig. 1 Regulation of T-lymphocyte proliferation in a, normal donor and b, LAD patient by anti-CD11a (CLB-LFA-1/2, squares) and anti-CD18 (CLB-LFA-1/1 triangles) antibodies. Also shown is the control (●) without added antibody.

**Methods.** Lymphocytes were isolated from PMBC by counterflow centrifugation elutriation. Residual monocytes in these lymphocyte suspensions were removed by a 2 h plastic-adherence step. The final suspension contained less than 1% monocytes. Immunofluorescence studies showed that LFA-1 expression on lymphocytes of the LAD patient used was <1% compared to the expression on normal donor lymphocytes<sup>7</sup>. Flat-bottomed wells (Nunc, Denmark) were coated overnight with graded quantities of protein-A-purified anti-CD3 antibody CLB-T3/3 diluted in phosphate-buffered saline (PBS). After 16 h incubation, the coated wells were washed three times with PBS. Lymphocytes were cultured for 4 d at 37°C in IMDM supplemented with 10% FCS and antibiotics (●). CLB-LFA-1/2 (■) and CLB-LFA-1/1 (▲) were added to the coated wells after dilution in culture medium to a final ascites dilution of 1:10<sup>3</sup>. Results are shown as the mean of triplicate cultures along the vertical axis.

\* To whom correspondence should be addressed.

**Table 1** The role of LFA-1 molecules in T-cell proliferation induced by soluble anti-CD3 antibody

| Peripheral blood mononuclear cells of an LAD patient do not proliferate in response to soluble anti-CD3 antibody: |    |                     |                     | Anti-LFA-1- $\alpha$ and - $\beta$ mAb inhibit T-cell proliferation induced by soluble anti-CD3 antibody: |                         |                         |
|-------------------------------------------------------------------------------------------------------------------|----|---------------------|---------------------|-----------------------------------------------------------------------------------------------------------|-------------------------|-------------------------|
| c.p.m.                                                                                                            |    |                     |                     | c.p.m.                                                                                                    |                         |                         |
| Anti-CD3                                                                                                          | MO | PBMC <sub>CON</sub> | PBMC <sub>LAD</sub> | Concentrations ( $\mu\text{g ml}^{-1}$ )                                                                  | CLB-LFA-1/2( $\alpha$ ) | CLB-LFA-1/1 ( $\beta$ ) |
| —                                                                                                                 | —  | 100                 | 100                 | 25                                                                                                        | 1,700                   | 300                     |
| —                                                                                                                 | +  | 100                 | 100                 | 6.25                                                                                                      | 2,700                   | 800                     |
| +                                                                                                                 | —  | 6,900               | 300                 | 1.5                                                                                                       | 6,000                   | 700                     |
| +                                                                                                                 | +  | 11,400              | 1200                | 0                                                                                                         | 8,900                   | 8,900                   |

Peripheral blood mononuclear cells of a healthy donor (PBMC<sub>CON</sub>) and of a LAD patient (PBMC<sub>LAD</sub>) were isolated by Ficoll-Isopaque ( $d = 1.079$ ) density centrifugation of heparinized blood.  $4 \times 10^4$  cells were cultured for 3 days in Iscove's modified Dulbecco's medium (IMDM) supplement with 10% fetal calf serum (FCS) and antibiotics. Stimulation was carried out with  $0.1 \mu\text{g ml}^{-1}$  CLB-T3/3 in the absence or presence of  $1 \times 10^4$  irradiated allogeneic monocytes (MO). During the last 4 h of culture,  $0.73 \text{ kBq } [^3\text{H}]$  thymidine ( $3.6 \times 10^9 \text{ Bq per mmol}$ ) was added to measure cell proliferation. Results are shown as the mean counts per minute (c.p.m.) of triplicate cultures. Graded quantities of affinity-purified anti-LFA-1 antibody were added at the start of the culture. A control antibody (CD45) showed no inhibition (data not shown).

formed clusters within 24 h of initiation of the culture, whereas the cells of the LAD patient showed no such tendency (not shown). Because these experiments showed that intercellular adhesion involving LFA-1 molecules is not necessary for the induction of T-cell proliferation in this system, we have used it to study the potential regulatory signals elicited through ligand binding to LFA-1  $\alpha$  and  $\beta$  chains in human T-cell activation.

In agreement with other reports<sup>16-18</sup>, addition of either anti-LFA-1- $\alpha$  (CLB-LFA-1/2; CD11a) or anti-LFA-1- $\beta$  (CLB-LFA-1/1; CD18) monoclonal antibody<sup>19</sup> to normal donor T cells stimulated with soluble anti-CD3 antibody resulted in strong inhibition of T-cell proliferation (Table 1). But when immobilized anti-CD3 antibody was used as stimulant, addition of CLB-LFA-1/2 strongly enhanced T-cell mitogenesis, whereas CLB-LFA-1/1 still inhibited T-cell proliferation (Fig. 1a). Inhibitory and stimulatory effects were both most apparent when suboptimal anti-CD3 stimulation was used (Fig. 1a), and when the antibodies were added during the first hour of culture (data not shown). No influence on T-cell proliferation occurred when these antibodies were added to cultures of T cells from the LAD patient (Fig. 1b). We tested a CD11a and CD18 panel<sup>19</sup> and observed a similar functional distinction between anti- $\alpha$  and anti- $\beta$  antibodies (Table 2). Five out of six anti-CD18 antibodies tested markedly suppressed T-cell proliferation, whereas all anti-CD11a antibodies enhanced anti-CD3-induced mitogenesis to a varying extent. The enhancement of proliferation was most evident for CLB-LFA-1/2. Therefore the epitope specificity and/or avidity of the anti-LFA-1- $\alpha$  antibodies probably contributed to the costimulatory effect. The specificity of the costimulatory effect of the anti- $\alpha$  antibody was tested using a large panel of monoclonal antibodies directed against T-lymphocyte membrane molecules in this system. Monoclonal antibodies directed against CD2, CD3, CD4, CD6 and CD25 blocked T-cell proliferation, whereas anti-CD7 and anti-CD8 antibodies did not alter the response. By comparison with CLB-LFA-1/2, there was little enhancement of proliferation after addition of anti-CD5 and anti-CD27 antibodies (data not shown).

The proliferation inhibition by CLB-LFA-1/1 and enhancement by CLB-LFA-1/2 is associated with changes in the production of interleukin 2 (IL-2) and with altered expression of the IL-2 receptor (Fig. 2). After 40 h of culture, T-cell suspensions costimulated with CLB-LFA-1/2 contained twice the number of CD25 antigen-positive cells and produced up to ten times more IL-2 compared with control cultures. Moreover, the anti-CD11a-enhanced proliferation could be largely blocked by the addition of anti-CD25 mAb (not shown). In contrast, suspensions stimulated in the presence of CLB-LFA-1/1 produced little IL-2 and contained few CD25<sup>+</sup> cells compared with the control cultures (Fig. 2). Thus, the mechanisms underlying pro-

**Table 2** Influence of a panel<sup>19</sup> of anti-CD11a ( $\alpha$ ) and anti-CD18 ( $\beta$ ) antibodies on T-cell proliferation induced through immobilized anti-CD3 antibodies

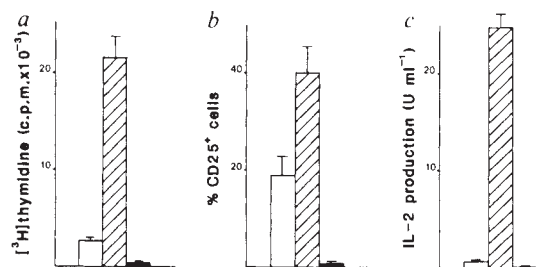
| Anti-CD11a  |      | Anti-CD18   |      |
|-------------|------|-------------|------|
| MHM 24      | 133% | TS1/18      | 101% |
| TS 1/22     | 150% | 1B4         | 4%   |
| CLB-LFA-1/2 | 264% | CLB-LFA-1/1 | 1%   |
| 25.3.1      | 145% | M 232       | 8%   |
| CIMT        | 135% | 60.3        | 11%  |
| 2F12        | 153% | MHM 23      | 2%   |
| TS 1/12     | 132% |             |      |
| TS 2/6      | 193% |             |      |
| 25.5.2      | 149% |             |      |

Lymphocytes were stimulated as described in the legend to Fig. 1 with  $0.08 \mu\text{g ml}^{-1}$  coated CLB-T3/3. Before the initiation of the culture, anti-CD11a and anti-CD18 antibodies were added to the culture wells. The final ascites dilution of antibody was  $1:10^3$ . Proliferation upon stimulation with  $0.08 \mu\text{g ml}^{-1}$  CLB-T3/3 was 6,400 c.p.m. on day 4. The change in proliferation is shown as: c.p.m. in the presence of antibody/6,400  $\times 100\%$ .

liferation inhibition by anti-LFA-1- $\beta$  and proliferation enhancement by anti-LFA-1- $\alpha$  antibodies probably mediate their effects by the IL-2 production/IL-2 receptor pathway.

Thus, using a T-cell activation model independent of cellular adhesion mediated by LFA-1, we have shown that antibodies directed against the LFA-1  $\beta$  subunit generally inhibit T-cell proliferation and anti-LFA-1- $\alpha$ -chain antibodies enhance T-cell growth to a varying extent. This latter observation is in agreement with murine studies reporting agonistic effects on lymphocyte activation of anti-LFA-1- $\alpha$  antibody (refs 22-23). The biological significance of the functional distinction between anti-LFA-1- $\alpha$  and anti-LFA-1- $\beta$  antibodies is not clear at present. So far, only one ligand for the LFA-1 molecule (ICAM-1) has been identified, although more than one counterstructure could exist<sup>3</sup>. Distinct LFA-1-ligand molecules may vary in the degree of binding to either  $\alpha$  or  $\beta$  chains and thus contribute to the regulatory signals on cell activation that accompany intercellular adhesion. Alternatively, the  $\alpha$  and  $\beta$  subunits of the heterodimeric LFA-1 molecule may have distinct biological functions<sup>24</sup>. By analogy with the CD3/TCR complex<sup>25</sup>, the  $\alpha$  chain could act as the ligand-binding molecule, with the  $\beta$  chain functioning as the signal-transducing element of the LFA-1 molecule without being involved in ligand binding; binding of the ligand to the  $\alpha$  chain would then be associated with the generation of agonistic signals transduced through the  $\beta$  chain to the intracellular compartment. In contrast, binding of non-





**Fig. 2** Change in IL-2-receptor expression and IL-2 production on cocultivation with CLB-LFA-1/1 (filled), CLB-LFA-1/2 (cross-hatched). Control without added antibody is shown blank.

**Methods.** Lymphocytes were cultured as described in the legend to Fig. 1 using  $0.08 \mu\text{g ml}^{-1}$  coated CLB-T3/3. After 40 h of culture, cells and supernatant were collected. IL-2-receptor expression was measured by fluorescence-activated cell sorting using FITC-labelled CLB-IL-2R/1 (b) (ref. 20). The percentage of CD25<sup>+</sup> cells were calculated after correction for nonspecific binding of the conjugate to unstimulated cells. IL-2 production was quantitated using the method described by Gillis *et al.*<sup>21</sup> (c). In parallel cultures, proliferation was measured on day 4 (a). Results are shown as the mean  $\pm$  s.d. of three separate experiments.

physiological ligands to the  $\beta$  chain could induce antagonistic signals generally to inhibit T-cell activation.

Cellular adhesion molecules are thought to determine migration, homing and positioning of cells in embryogenesis and histogenesis<sup>26-28</sup>. LFA-1 belongs to a large supergene family of adhesion molecules present on different cell types of different species<sup>26,29</sup>, so its signal-modifying properties may have a more general biological relevance. Thus, binding of ligands present on encountered cells or in extracellular matrices is important not only for cell positioning, but could possibly also regulate cell growth and differentiation.

We thank A. J. C. Petit, P. van Mourik and F. G. Terpstra for technical assistance and Drs C. J. M. Melief, W. P. Zeijlmeier, G. A. Van Severter and C. J. Lucas for reviewing the manuscript.

Received 11 March; accepted 27 May 1988.

- Springer, T. A., Dustin, M. L., Kishimoto, T. K. & Marlin, S. D. *A. Rev. Immun.* **5**, 223-252 (1987).
- Martz, E. *Human Immun.* **18**, 3-37 (1987).
- Rothlein, R., Dustin, M. L., Marlin, S. D. & Springer, T. A. *J. Immun.* **137**, 1270-1274 (1986).
- Makgoba, M. W. *et al. Nature* **331**, 86-88 (1988).
- Marlin, S. D. & Springer, T. A. *Cell* **51**, 813-819 (1987).
- Anderson, D. C. *et al. Fedn. Proc.* **44**, 2671-2677 (1985).
- Miedema, F. *et al. J. Immun.* **134**, 3075-3081 (1985).
- Hunig, T. R., Tiefenthaler, G., Meyer zum Büschenfelde, K. H. & Meuer, S. C. *Nature* **326**, 298-300 (1987).
- Bank, I. & Chess, L. *J. exp. Med.* **162**, 1294-1303 (1985).
- Fleisher, B., Schrezenmeier, H. & Wagner, H. *J. Immun.* **136**, 1625-1628 (1986).
- Van Severter, G. A., Van Lier, R. A. W., Spits, H., Ivanyi, P. & Melief, C. J. M. *Eur. J. Immun.* **16**, 1363-1371 (1986).
- Van Wauwe, J. P., De May, J. R. & Goossens, J. G. *J. Immun.* **124**, 2708-2713 (1980).
- Tax, W. J. M., Hermes, F. F. M., Willems, R. W., Capel, P. J. A. & Koene, R. A. P. *J. Immun.* **133**, 1185-1189 (1984).
- Van Lier, R. A. W., De Groot, E. R., Rinnoy Kan, E. A., Zeijlmeier, W. P. & Aarden, L. A. in *Tissue Typing III* (eds McMichael, A. J. *et al.*) 181 (Oxford University Press, 1987).
- Geppert, T. D. & Lipsky, P. E. *J. Immun.* **138**, 1660-1666 (1987).
- Sanchez-Madrid, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **70**, 7489-7493 (1982).
- Miedema, F. *et al. Eur. J. Immun.* **14**, 518-528 (1984).
- Dougworth, D. W., Gotch, F. M., Hildreth, J. E. K., Morris, A. & McMichael, A. J. *Eur. J. Immun.* **15**, 888-892 (1985).
- Waldman, H. *et al. in Tissue Typing III* (eds McMichael, A. J. *et al.*) 829 (Oxford University Press, 1987).
- Van Lier, R. A. W., Brouwer, M. & Aarden, L. A. *Eur. J. Immun.* **18**, 167-172 (1988).
- Gillis, S. *et al. J. Immun.* **120**, 2027-2032 (1978).
- Mishra, G. C. *et al. J. Immun.* **137**, 1590-1598 (1986).
- Pircher, H. *et al. Eur. J. Immun.* **16**, 172-181 (1986).
- Ross, G. D. & Medof, M. D. *Adv. Immun.* **37**, 217-267 (1985).
- Tsoukas, C. D., Valentine, M., Lotz, M., Vaughan, J. H. & Carson, D. A. *Immun. Today* **5**, 311-313 (1984).
- Ruoslahti, E. & Pierschbacher, M. D. *Science* **238**, 491-497 (1987).
- Gallatin, M. *et al. Cell* **44**, 673-680 (1986).
- Cunningham, B. A. *TIBS* **11**, 423-426 (1986).
- Kishimoto, T. K., O'Connor, K., Lee, A., Roberts, T. M. & Springer, T. A. *Cell* **48**, 681-690 (1987).

## Transgenes as probes for active chromosomal domains in mouse development

Nicholas D. Allen\*, David G. Cran\*, Sheila C. Barton\*, Simon Hettle†, Wolf Reik\* & M. Azim Surani\*

\* Department of Molecular Embryology, Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, UK

† University of Glasgow, Department of Veterinary Pathology, Veterinary School, Bearsden Road, Glasgow G61 1QA, UK

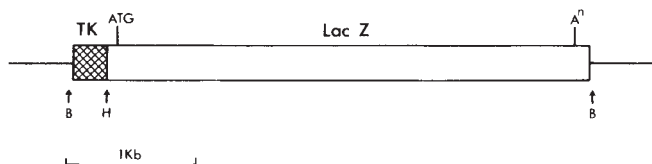
Embryonic development entails a well defined temporal and spatial programme of gene expression, which may be influenced by active chromosomal domains. These chromosomal domains can be detected using transgenes which integrate randomly throughout the genome, as their expression can be affected by chromosomal position<sup>1-3</sup>. Position effects are probably exerted most strongly on transgenes that do not contain strong promoters, enhancers or other modulating sequences. Here we have systematically explored position effects using a transgene with the weak herpes-simplex-virus thymidine-kinase promoter, linked to the readily visualized *lacZ* indicator gene (HSV-TK-*lacZ*). Each transgenic fetus with detectable expression displayed a unique *lacZ* staining pattern. Thus expression of this construct is apparently dictated entirely by its chromosomal position, without any construct specificity. Furthermore the transgene is faithfully transmitted to subsequent generations, allowing for systematic mapping of changes in expression during development and in adult life. These results demonstrate that transgenes can indeed be powerful tools to probe the genome for active chromosomal regions, with the potential for identifying endogenous genes involved in organogenesis and pattern formation.

The gene construct, pHFBG2 (Fig. 1), has a weak promoter which simply provides the necessary immediate 5' flanking sequences<sup>4</sup> to permit effective expression of the *lacZ* gene<sup>5</sup>. HSV-TK has no known developmental or tissue specific regulation and the HSV-TK gene can be expressed as a transgene in both preimplantation and midgestation embryos<sup>6,7</sup>. More importantly, modulation of TK gene activity by chromosomal position has been demonstrated in cell culture<sup>8</sup>. The *lacZ* gene is an effective *in situ* marker gene which has been used in cell lineage analysis with retroviral vectors<sup>9,10</sup>, as well as in transgenic mice<sup>11</sup>, without apparent detrimental effects on cells expressing the *lacZ*-encoded enzyme  $\beta$ -galactosidase. In addition, clonal stability of expression has been observed, at least in tissue culture cells (J-F Nicolas, personal communication).

HSV-TK-*lacZ* DNA ( $2\text{ng } \mu\text{l}^{-1}$ ) was injected into the male pronucleus of fertilized mouse eggs (C57BL/6  $\times$  CBA  $\delta$ ) F<sub>1</sub>  $\times$  C57BL/6  $\delta$ <sup>12</sup>. High levels of transient expression had previously been observed in ~50% of one and two-cell embryos (data not shown). Transgenic fetuses were analysed for *lacZ* expression on day 11 of gestation, a key time during development which is associated with major changes in organogenesis as well as growth, and therefore presumably has a large number of developmentally active genes and chromosomal domains.

For each fetus, the extra-embryonic membranes (trophoblast and yolk sac) were used to prepare DNA, while the embryos were fixed and stained (see legend to Fig. 2). Of the 200 embryos analysed, 52 were positive for DNA integration and 11 of these were found to have distinct expression patterns (Table 1). The transgene copy number ranged from ~5-500 copies in tandem repeats constituting domains of between 20-2,000 kilobases (kb), with no evidence of independent integration events (see below).

Two broad categories of expression were apparent. In five embryos (numbers 1, 7, 75, 157, 177) the patterns observed were complex with expression in two or more disparate tissues. Embryo 7 (Fig. 2a) expressed the transgene in two tissues; the



**Fig. 1** Gene construct pHFBG2. The 3.69 kilobase (kb) HSV-TK-*lacZ* fragment was excised from its vector sequences by restriction with *Bam*HI (B). It contains 240 base pairs of the HSV-TK 5' promoter region<sup>4</sup> (■), ligated to the *lacZ* structural gene derived from pCH110<sup>5</sup> (□) at the *Hind*III site (H). The polyadenylation signal is derived from simian virus 40.

pattern was segmentally repeated in association with every somite and in addition there was staining in the forelimb bud. Embryo 1 (Fig. 2b) was far more complex, with expression in seemingly unrelated tissues (brain, neural tube, skin, heart and somites). In the other six embryos expression was confined to specific tissues, namely the heart, skin, neural tube, spinal ganglia and mesonephric ducts. Expression in the latter two embryos is shown in Fig. 2c, d and e, and 2f and g respectively. The expression can be more precisely characterized in tissue sections; under dark-field microscopy the stain is an intense red colour (Fig. 2e, g). In addition to cell-type specificity, some embryos also exhibited spatially specific expression. Thus in embryo 75, staining was restricted to spinal ganglia at somite segments in only the lumbar region and to ganglia in the brain. A similar pattern of discrete blue spots within a few somites was seen in embryo 59, but this staining was cranial to that seen in embryo 75 (Fig. 2c). One might have expected to observe low but ubiquitous expression from the HSV-TK promoter and constitutive expression has indeed been observed in fetuses infected with retroviral vectors carrying a gene under the control of an internal TK promoter<sup>13</sup>. Whether this reflects preferential insertion of these vectors into constitutively 'open' chromatin<sup>14</sup> or a function of the retroviral genome is not clear. In our case, however, low levels of expression may fall below the sensitivity of the assay.

The results show that in each embryo in which expression was detected, the pattern of expression showed unique tissue or spatial specificity. In addition, expression was observed in derivatives of all the three definitive germ layers (see Table 1). These findings indicate that the expression of this transgene is indeed dependent on the chromosomal site of insertion. Nevertheless, modulation of position effects by the transgenome itself is possible. The use of cloned DNA rather than retroviruses results in variable multiple copy insertions (Table 1) mostly in head to tail arrangement<sup>1</sup>. The creation of novel specificities could arise from the juxtaposition of unrelated sequences in a hybrid transgene and their novel molecular arrangement in tandem repeats. Such specificities have previously been reported to be highly reproducible, however<sup>15,16</sup>. It is also possible that different copies within a tandem repeat might respond differently to a position effect. Indeed, different levels of expression of

individual copies of such tandems have been reported, but the developmental profile and coregulation of the copies under study was the same<sup>17</sup>.

In a few studies in mice, anomalous transgene expression and methylation patterns have been attributed to position effects<sup>2,3,18-20</sup>, but in most cases a systematic molecular analysis of these position effects has not been done<sup>2,3,18,19</sup>. A molecular analysis has been made of integration sites in transformed cell cultures, where it is possible to select for position-dependent expression. On every occasion that this has been done, sequences with the properties of enhancers have been found. This applies to both retroviruses and DNA sequences with variable copy numbers in tandem<sup>21-25</sup>.

At least two kinds of position effect can operate on the transgene. Expression in specific tissues may indicate *cis*-acting effects from nearby enhancers or other tissue-specific regulatory units. Alternatively, the transgene could have inserted directly into a cellular gene with the consequence of transcriptional read-through into the *lacZ* gene. The second category of expression in seemingly disparate tissues may arise because the transgene resides within a large and active chromosomal domain in which endogenous genes are also expressed in different tissues, possibly to effect coordination of gene expression throughout the embryo. An example of such a domain which affects spatial specificity, while remaining developmentally constant, is the  $\beta$ -globin region that contains several genes and is flanked by so called 'locus activation regions'<sup>26,27</sup>.

It is possible that some of the complex expression patterns observed are due to DNA integration mosaicism. The number of embryos exhibiting disparate and diffuse patterns greatly exceeds the expected frequency of 20-30% of mosaicism in transgenic mice generated by pronuclear microinjection of DNA, however<sup>28</sup>. Moreover, because all the staining patterns that we observed were identical in symmetrical structures, this essentially rules out a major contribution of mosaicism in this study. Similar findings to those described here have recently been reported by O'Kane and Gehring<sup>29</sup> who employed a very similar strategy of using position-dependent transgenes for identifying developmentally-regulated chromatin regions in *Drosophila*.

We have now produced 20 transgenic lines using pHFBG2. Five of these show unique expression patterns during development, while in the remaining 15 lines no expression has so far been observed over a number of developmental stages tested. These studies further support the view that the transgene expression is determined by the site of integration. The faithful transmission of the transgene, and hence the expression pattern, to subsequent generations presents a unique opportunity to study the temporal and spatial changes of gene expression.

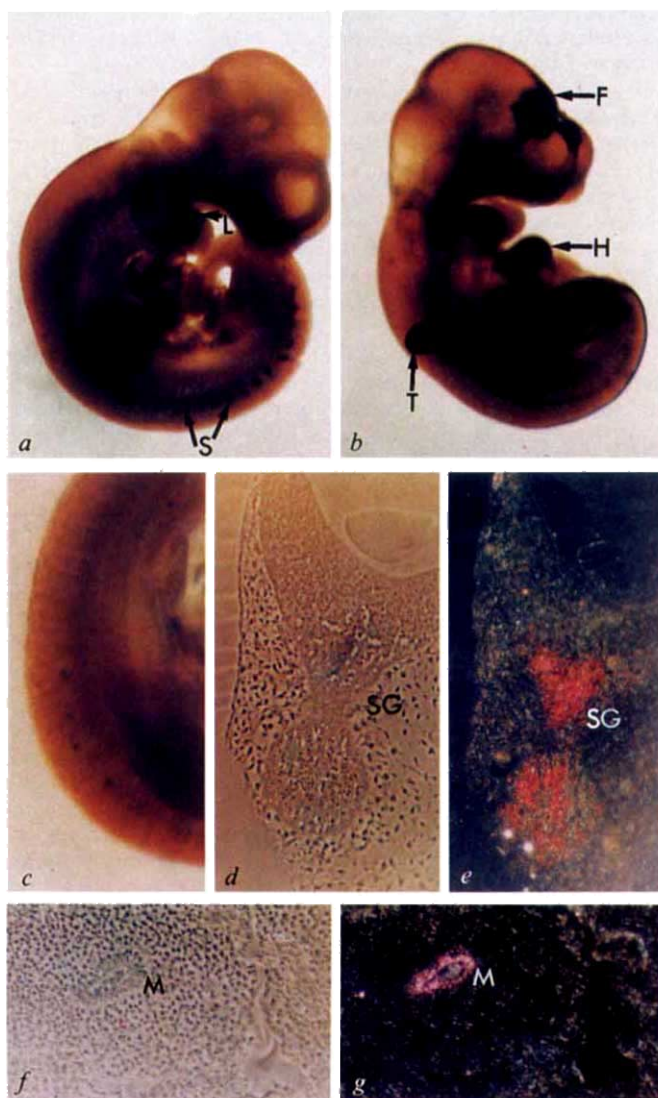
One of the transgenic lines under study (TKZ710) uniquely expresses the transgene in the brain, including the adult brain. The expression of this transgene has been mapped at different developmental stages. Expression is first detected on early day 11 of gestation in a small group of cells scattered in the hind

**Table 1** Summary of *lacZ* expression patterns observed in transgenic fetuses

| Embryo number                     |                   | 1* | 7*   | 22* | 23*  | 49*   | 59* | 75† | 110* | 157† | 177† | 200† |
|-----------------------------------|-------------------|----|------|-----|------|-------|-----|-----|------|------|------|------|
| Approximate transgene copy number |                   | 25 | 5-10 | 100 | 5-10 | 10-15 | 50  | 10  | ND   | 5-10 | 15   | 20   |
| Ectoderm-derived tissues          | Brain             | +  |      |     |      |       |     | +   |      | +    |      |      |
|                                   | Neural tube       | +  |      | +   |      |       |     |     |      | +    | +    |      |
|                                   | Spinal ganglia    |    |      |     |      |       | +   | +   |      |      |      |      |
|                                   | Forelimb bud      |    | +    |     |      |       |     |     | +    |      |      |      |
|                                   | Skin              | +  |      |     | +    |       |     |     |      |      |      |      |
| Mesoderm-derived tissues          | Heart             | +  |      |     |      | +     |     |     |      |      |      |      |
|                                   | Somites           | +  | +    |     |      |       |     |     |      |      | +    |      |
|                                   | Mesonephric ducts |    |      |     |      |       |     |     |      |      |      | +    |
| Endoderm-derived tissues          | Lung bud          |    |      |     |      |       |     |     |      | +    |      |      |

\* Expression patterns defined by embryo dissection. † Expression patterns defined in tissue sections.

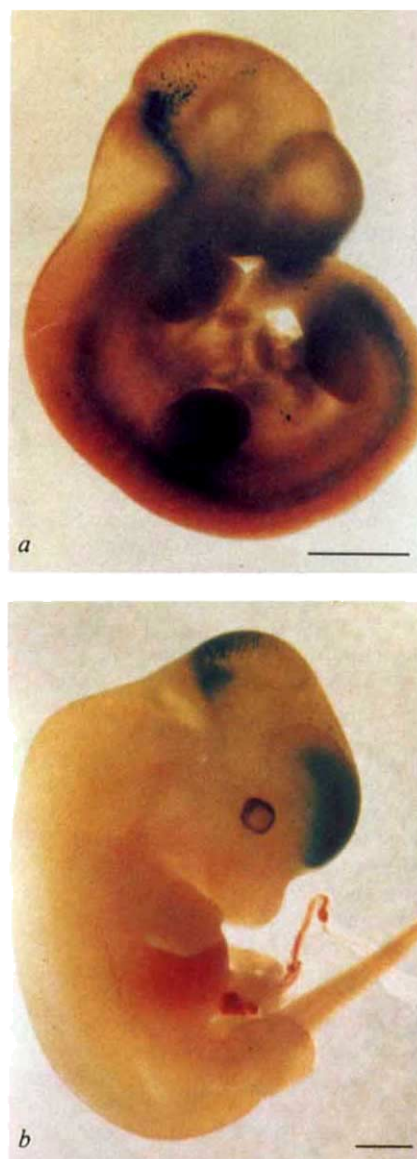




**Fig. 2** Histochemical staining of HSV-TK-*lacZ* transgenic embryos for  $\beta$ -galactosidase. The blue stain is readily discernible in whole embryos *a*, *b* and *c*. Staining can be precisely localized in tissue sections by dark-field microscopy, (*e*, *g*) and compared with the phase-contrast image (*d*, *f*). Embryo 7, shown in *a*, expressed the transgene in the somites (S) and forelimb buds (L). Embryo 1 (*b*) expressed the transgene in its forebrain (F), heart (H) and tail (T). Sections of embryo 75 (*c*) showed the discrete blue spots to be spinal ganglia (SG) (*d*, *e*). Expression in embryo 200 was restricted to the mesonephric ducts (M) (*f*, *g*).

**Methods.** Embryos were recovered on day 11 of gestation and fixed for 30 min at 4 °C in 1% formaldehyde, 0.2% glutaraldehyde and 0.02% NP40 in phosphate buffer saline (PBS). Fixed embryos were washed 3 $\times$  in PBS and incubated in stain solution (5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 1 mg ml<sup>-1</sup> X-gal in PBS) overnight at 30 °C<sup>9</sup>. Whole embryos were examined under a dissecting microscope and photographed. Embryos for sectioning were embedded in LR white acrylic resin, sectioned at 2–4  $\mu$ m and viewed either with phase-contrast or dark-field optics.

brain, together with a transient expression in cells within the neural tube and somites (Fig. 3*a*). By day 13 of gestation (Fig. 3*b*) the transgene is expressed diffusely throughout telencephalic structures (cortex, anterior olfactory nucleus, hippocampus, induseum griseum) and in a small population of cells in the roof of the IVth ventricle which are destined to become the cerebellum. By day 19 of gestation the telencephalic expression becomes focused to frontal cortex and temporal cortex (hip-



**Fig. 3** Expression of *lacZ* in embryos from the transgenic line TKZ710. Expression is first seen on day 11 (*a*) of gestation in the hind brain and within the neural tube and somites. By day 13 of gestation (*b*), expression has become restricted to the brain (see text for details). Embryos were fixed and stained as described in the legend to Fig. 2. Scale bars, 1 mm.

pocampus, entorhinal and piriform cortices) and is substantially concentrated in the olfactory bulbs and anterior olfactory nucleus. In the postnatal period on day 5, expression continues to disappear from the cortex and is concentrated in areas concerned with olfactory processing, but also appears in the pontine nuclei which project to the Purkinje cells of the cerebellum. This expression is then maintained in adults. In the olfactory bulbs of the adult mouse expression of the transgene is localized in periglomerular cells, and in the hippocampus is restricted to pyramidal cells of CA2, but still appears in the hippocampal rudiment, the induseum griseum. Also in the adult, the transgene expression appears for the first time in a diencephalic structure, the mammillary nucleus, a structure to which the hippocampus projects. The transgene seems to be expressed in neural elements rather than glia (N.D.A. and E. B. Keverne, in preparation).

These findings demonstrate changes in transgene expression in the developing brain that are restricted to a wide range of

telencephalic structures. In the postnatal brain these changes in expression become more focused to specific cell lines, many of which have olfactory processing in common. It is particularly interesting that this transgene continues to be expressed in the postnatal and even the adult brain. It has provided an invaluable tool for mapping telencephalic development in a structural way and in addition, of probing the functional development of the olfactory system.

In this and other transgenic strains being studied, in which expression is detected in other developing systems, such as gonads and mammary glands, the transgene serves as a genetic marker from which a full molecular analysis of the integration sites can be made, which will lead to the identification of endogenous genes or other regulatory sequences responsible for

the transgene position-effects observed. In conjunction with the detailed mapping of expression over developmental time as demonstrated here, this approach of using position-dependent transgenes is an immensely powerful way of systematically analysing mammalian development in a non-subjective manner.

We thank Barry Keverne, Sarah Howlett, Michael Norris, Reinald Fundele and Mu-Chiou Huang for discussion of the manuscript. We also thank Linda Notton for typing the manuscript. N.D.A. is supported by an AFRC studentship. W.R. is a fellow of the Lister Institute of Preventive Medicine. The work was supported in part by the MRC Laboratory of Molecular Biology; we thank our colleagues at the LMB, especially Michael Neuberger and Terry Rabbitts, for valuable help and encouragement.

Received 5 May; accepted 27 May 1988.

- Palmiter, R. D. & Brinster, R. L. *A. Rev. Genet.* **20**, 465-499 (1986).
- Jaenisch, R. *et al. Cell* **24**, 510-529 (1981).
- Lacy, E., Roberts, S., Evans, E. P., Burtenshaw, M. D. & Costantini, F. D. *Cell* **34**, 343-358 (1983).
- Wagner, M. J., Sharp, J. A. & Summers, W. C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1441-1445 (1981).
- Hall, C., Jacob, P., Ringold, G. & Frank, L. *J. molec. appl. Genet.* **2**, 101-109 (1983).
- Wagner, E. F., Stewart, T. A. & Mintz, B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5016-5020 (1981).
- Brinster, R. L., Chen, H. Y., Warren, R., Sarthy, A. & Palmiter, R. D. *Nature* **296**, 39-42 (1982).
- Butner, K. & Lo, C. W. *Molec. Cell Biol.* **6**, 4440-4449 (1986).
- Sanes, J. R., Rubenstein, J. L. R. & Nicolas, J.-F. *EMBO J.* **5**, 3133-3142 (1986).
- Turner, D. L. & Cepko, C. L. *Nature* **328**, 131-136 (1987).
- Goring, D. R., Rossant, J., Clapoff, S., Breitman, M. L. & Tsui, L. C. *Science* **235**, 456-457 (1987).
- Allen, N. D., Barton, S. C., Surani, M. A. H. & Reik, W. in *Mammalian Development—A Practical Approach* (ed. Monk, M.) 217-233 (IRL Press, Oxford, 1987).

- Stewart, C. L., Schuetz, S., Vanek, M. & Wagner, E. F. *EMBO J.* **6**, 383-388 (1987).
- Rohdewold, H., Weiher, H., Reik, W., Jaenisch, R. & Breindl, M. *J. Virol.* **61**, 336-343 (1987).
- Swanson, L. W. *et al. Nature* **317**, 363-366 (1985).
- Low, M. J. *et al. Science* **231**, 1002-1024 (1987).
- Das, O. P. & Massing, J. W. *Molec. Cell Biol.* **7**, 4490-4497 (1987).
- Soriano, P., Cone, R., Mulligan, R. & Jaenisch, R. *Science* **234**, 1409-1413 (1986).
- Shani, M. *Molec. Cell Biol.* **6**, 2624-2631 (1986).
- Jahner, D. & Jaenisch, R. *Molec. Cell Biol.* **5**, 2212-2220 (1985).
- Sorge, J., Cutting, A., Erdman, V. & Gautsch, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6627-6631 (1984).
- Barklis, E., Mulligan, R. & Jaenisch, R. *Cell* **47**, 391-399 (1986).
- Taketo, M. & Tanaka, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3748-3752 (1987).
- Hamada, H. *Molec. Cell Biol.* **6**, 4179-4184 (1986).
- Hamada, H. *Molec. Cell Biol.* **6**, 4185-4194 (1986).
- Sorge, W. C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. & Groudine, M. *Nucleic Acid Res.* **15**, 10159-10177 (1987).
- Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. *Cell* **51**, 975-985 (1987).
- Wilkie, T. M., Brinster, R. L. & Palmiter, R. D. *Dev. Biol.* **118**, 9-18 (1986).
- O'Kane, C. J. & Gehring, W. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9123-9127 (1987).

## Expression of the $\gamma$ - $\delta$ T-cell receptor on intestinal CD8<sup>+</sup> intraepithelial lymphocytes

Thomas Goodman & Leo Lefrançois

Department of Cell Biology, The Upjohn Company, Kalamazoo, Michigan 49001, USA

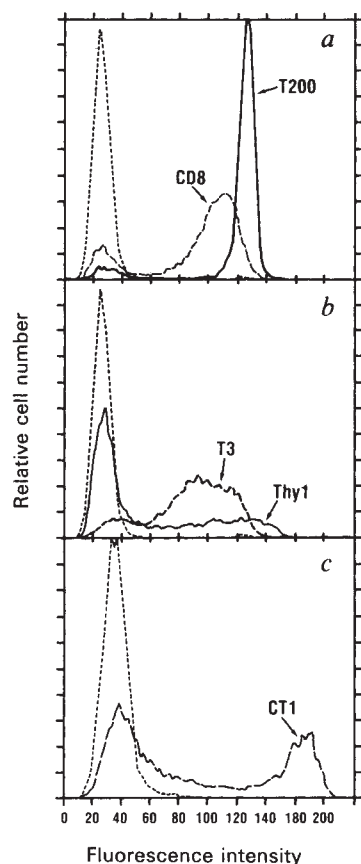
The vast majority of mature T lymphocytes in the peripheral blood and lymphoid organs use the CD3-associated  $\alpha$ ,  $\beta$  T-cell receptor (TCR) heterodimer for antigen recognition<sup>1-3</sup>. A second class of TCRs consists of disulphide-linked  $\gamma$  and  $\delta$  proteins that are also CD3-associated<sup>4-7</sup>. A subset of early CD3<sup>+</sup> fetal and adult CD4<sup>-</sup> 8<sup>-</sup> thymocytes express  $\gamma$ ,  $\delta$  TCRs before  $\alpha$ ,  $\beta$  TCRs are detectable<sup>7-10</sup>. In addition, a minor (1-5%) subpopulation of peripheral T lymphocytes<sup>4,11,12</sup>, and some spleen cells from nude mice<sup>13</sup> express  $\gamma$ ,  $\delta$  TCRs. Notably, dendritic epidermal cells have also been shown to express  $\gamma$ ,  $\delta$  TCRs<sup>14,15</sup>. All of these populations lack CD4 and CD8 molecules. We now report that most mature T cells residing in the murine intestinal epithelium express CD3-associated TCRs composed of  $\gamma$ -chains disulphide-linked to a protein resembling the  $\delta$ -chain. The striking feature of these intraepithelial lymphocytes (IEL) was that they were exclusively CD4<sup>-</sup> 8<sup>+</sup>. In addition, approximately half of CD3-bearing IEL lacked detectable Thy-1 on the cell surface, which is unprecedented for murine T cells. In contrast to other CD8<sup>+</sup> peripheral T cells, freshly isolated IEL could be induced to display cytolytic activity by engaging the CD3 molecule, indicating that activation had occurred *in vivo*. Thus, CD8<sup>+</sup> IEL are a phenotypically diverse and anatomically restricted population of lymphocytes that use  $\gamma$ -chain containing heterodimers for antigen recognition.

Intestinal cells were obtained from C57BL/6J mice by dissection of small intestines, mechanical disruption of the tissue fragments and purification on Percoll gradients. As we reported previously, this cell population is enriched in IEL (50-70%) but contains epithelial cells<sup>16</sup>. Accordingly, IEL were routinely

analysed by fluorescence flow cytometry for expression of the T200 antigen (CD45), which is found only on cells of haematopoietic lineage. Of T200<sup>+</sup> IEL, 85% were CD4<sup>+</sup> 8<sup>+</sup>, 85-90% were CD3<sup>+</sup>, 40% were Thy1<sup>+</sup> and 50% were CT1<sup>+</sup> (Fig. 1). The remaining T200<sup>+</sup> IEL contained <5% CD4<sup>+</sup> 8<sup>-</sup> T cells and 7-10% sIg<sup>+</sup> cells (data not shown). These values are consistent with our previous results from two-colour fluorescence analysis<sup>16</sup> and with the results of others<sup>17,18</sup>. These results therefore define a novel T-cell population that is CD3<sup>+</sup> 8<sup>+</sup> and Thy-1<sup>-</sup>. In addition, IEL are the only peripheral T cells expressing the T200-associated carbohydrate antigen, CT1 (refs. 16, 19). The CT1 determinant is present on the CD45 molecule of fetal thymocyte subsets and CTL clones grown *in vitro*<sup>16,19</sup>, and on CD45 of IEL (Lefrançois and Goodman, unpublished results). The functional significance of the CT1 determinant is, as yet, unknown.

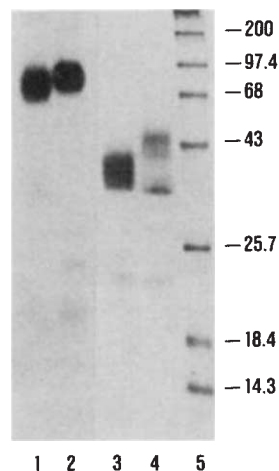
To determine the protein composition of the CD3-associated TCR of IEL, radioiodinated cells were subjected to immunoprecipitation with a hamster monoclonal antibody (mAb), 145-2C11, which is specific for the CD3- $\epsilon$  chain<sup>20</sup>. For comparison, the CD3<sup>+</sup> 4<sup>+</sup> 8<sup>+</sup>, Thy1<sup>+</sup> cytotoxic T cell (CTL) clone OE4 was used. The TCR of OE4 is recognized by the V $\beta$ <sub>8</sub>-reactive monoclonal antibody F23.1, and is composed of an  $\alpha$ ,  $\beta$ -heterodimer<sup>21,22</sup>. Anti-CD3 immunoprecipitation of digitonin lysates (digitonin does not disrupt the CD3-TCR association) of OE4 and analysis by non-reducing SDS-polyacrylamide gel electrophoresis (PAGE) yielded a major protein of apparent relative molecular mass ( $M_r$ ) 80,000 (80K; Fig. 2, lane 2). Analysis of similarly treated IEL revealed a labelled species of  $M_r$  78K (Fig. 2, lane 1). More striking differences were observed, however, when disulphide bonds of immunoprecipitated proteins were reduced. Whereas immunoprecipitates from OE4 cell extracts contained labelled proteins in the  $M_r$  range of 42K-44K, typical of  $\alpha$ ,  $\beta$  TCR chains (lane 3), IEL extracts contained two distinct proteins of  $M_r$  34K and 46K (lane 4), indicating that both TCRs were disulphide linked dimers. Several proteins of lower apparent  $M_r$  were also evident in precipitates from both cell types and represent subunits of the CD3 complex.





**Fig. 1** Fluorescence flow cytometry analysis of IEL. IEL were isolated as previously described except that 1 mM dithioerythritol (DTE) was used in place of EDTA<sup>34</sup>. Briefly, after removal of Peyer's patches, small intestinal fragments were stirred at 37 °C in the presence of 1 mM DTE twice for 20 min. Intestinal pieces were removed and the resulting cell population, containing epithelial cells and lymphocytes, was centrifuged through a stepwise 44% and 67.5% Percoll gradient. The cells at the 44%–67.5% interface were collected and washed before use. The cells were reacted with I3/2.3 (anti-T200), 3.168 (anti-CD8), T24 (anti-Thy-1), 145-2C11 (a hamster mAb specific for the CD3  $\epsilon$  chain<sup>20</sup>) or with CT1. In the case of I3/2.3, antibody binding was detected by subsequent reaction with a fluorescein isothiocyanate (FITC)-conjugated goat anti-rat immunoglobulin reagent (Cappel). The anti-Thy-1, anti-CD3 and anti-CD8 reagents were directly fluoresceinated. The CT1 mAb was biotinylated and detected by avidin-FITC. The dashed line represents staining with a control mAb and the FITC-labelled secondary reagent. Relative fluorescent intensities of individual cells were measured with an Ortho Cytofluorograph model 50H. In this experiment gating using forward and right angle light scatter was used to include only the lymphocyte population based on staining with the anti-T200 mAb. The results are presented as fluorescence histograms with the relative number of cells on a linear scale plotted versus the relative fluorescence intensity on a four decade logarithmic scale, both in arbitrary units.

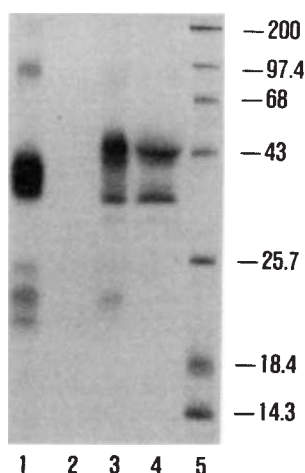
The  $M_r$ s of the proteins precipitated from IEL by the anti-CD3 monoclonal antibody correlate well with the reported  $M_r$  of the  $\gamma$ -chain (34–35K) and the  $\delta$ -chain (46K) from dendritic epidermal cells<sup>14,15</sup> and fetal and adult CD3<sup>+</sup>4<sup>+</sup>8<sup>+</sup> thymocytes<sup>7–10</sup> and are distinct from the  $M_r$ s of the  $\alpha$  and  $\beta$  chains precipitated from OE4 cells. Yet, it was possible that the observed CD3-associated proteins of IEL were members of an undescribed class of TCR or were differentially processed  $\alpha$  and  $\beta$  chains. In addition, as our labelled cell populations contained non-haematopoietic contaminants and a small percentage of CD8<sup>+</sup> cells we wished to confirm that the observed results were due to CD3-associated TCR of the CD8<sup>+</sup> IEL. Because reagents that react with murine surface  $\gamma$ ,  $\delta$  proteins are not yet available, immunoprecipitation analysis is required to identify these TCRs.



**Fig. 2** Immunoprecipitation of CD3-associated T-cell receptors from IEL. Cell surface proteins of IEL or of the cytotoxic T cell clone, OE4, were radioiodinated by the Iodogen method<sup>35</sup>. Precipitation by anti-CD3 was carried out essentially as described by Oettgen *et al.*<sup>36</sup> using anti-CD3 coupled to Sepharose 4B. Immunoprecipitates were analysed by SDS-PAGE using a 12.5% acrylamide gel in the absence (lanes 1, 2) or presence (lanes 3, 4) of 5% 2-mercaptoethanol. Lanes 1 and 4, IEL; lanes 2 and 3, OE4. Lane 5,  $M_r$  markers shown  $\times 10^{-3}$ .

We therefore purified CD8<sup>+</sup> IEL to 98% purity by panning on anti-CD8 mAb-coated petri dishes. Analysis by fluorescence flow cytometry of CD8<sup>+</sup> IEL indicated that 98% were CD3<sup>+</sup> and 37% were Thy 1<sup>+</sup> (data not shown). The cells were surface-iodinated and anti-CD3 precipitation was performed on digitonin lysates. A portion of the resulting CD3-TCR complexes was disrupted with SDS and reprecipitated with a rabbit anti-serum specific for the  $M_r$  34K–36K  $\gamma$  chains of the  $C_{\gamma 1-3}$  families<sup>7</sup>. A similar procedure was carried out with OE4 cells. The sequential immunoprecipitates from IEL and OE4 cells were analysed by SDS-PAGE under reducing conditions. Although the anti-CD3 mAb precipitated proteins of  $M_r$  42K–44K from OE4 lysates, the anti- $\gamma$  serum did not precipitate any detectable proteins (Fig. 3, lanes 1 & 2). Anti-CD3 precipitation from lysates of CD8<sup>+</sup> IEL yielded proteins of  $M_r$  34K and 46K and lower  $M_r$  CD3 proteins as observed in precipitates from unseparated intestinal cells (Fig. 3, lane 3). Immunoprecipitation with the anti- $\gamma$  chain serum yielded proteins of  $M_r$  34K and 46K after reduction but did not precipitate any of the CD3 proteins as expected. These results indicate that CD8<sup>+</sup> IEL predominantly express a  $\gamma$  chain of  $M_r$  34K, disulphide-linked to a  $M_r$  46K protein, probably the  $\delta$  chain. Two other protein species of  $M_r$  39K and 44K that appeared less prominent than the  $M_r$  34K and 46K species were observed in anti-CD3 precipitates from IEL. We have not as yet identified these proteins.

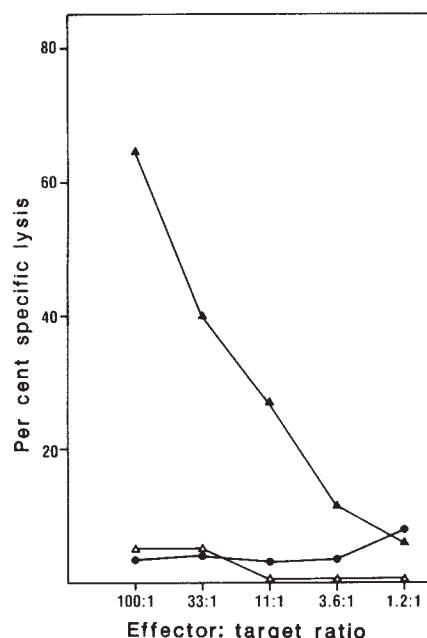
IEL have been reported to exhibit natural killer (NK) activity<sup>23,24</sup>, conventional cytotoxic T-cell activity<sup>25</sup>, and spontaneous cytotoxicity<sup>24</sup> (defined by lytic activity against P815 tumor targets). Other studies, including our own, demonstrate little cytotoxicity from freshly isolated IEL<sup>16,25,26</sup>. None of these studies analysed TCR use. To assess the inherent cytotoxic capabilities of CD8<sup>+</sup> IEL we used a redirected-lysis assay using anti-CD3 mAb<sup>27</sup>. The DBA/2 mastocytoma, P815, expresses an Fc receptor and when coated with anti-CD3 mAb can serve as a target for any CD3<sup>+</sup> cytotoxic T cell. When purified CD8<sup>+</sup> IEL were assayed for cytolytic activity against P815(H-2<sup>d</sup>), EL4(H-2<sup>b</sup>) and YAC-1(H-2<sup>a</sup>, NK sensitive) no lysis was observed (Fig. 4). When anti-CD3 mAb was included in the assay, however, P815 target cells were efficiently lysed by IEL but not by CD8<sup>+</sup> lymph node T cells. Based on the lack of target cell killing in the absence of anti-CD3 mAb, IEL did not appear to be alloreactive to H-2<sup>k,d</sup> and were not autoreactive to H-2<sup>b</sup>.



**Fig. 3** Immunoprecipitation of IEL T-cell receptor proteins by anti- $\gamma$  chain serum. IEL were further purified by panning on anti-CD8 coated petri dishes<sup>37</sup> and were shown to be 98% CD3<sup>+</sup>8<sup>+</sup> by fluorescence flow cytometry (not shown). CD8<sup>+</sup> IEL and OE4 cells were radioiodinated by the iodogen method and anti-CD3 precipitation was carried out on digitonin lysates as described<sup>36</sup>. For anti- $\gamma$  chain precipitation the anti-CD3 immunoprecipitates were disrupted by boiling for 3 min in the presence of 1% SDS. The samples were centrifuged and the supernatants were diluted five-fold with buffer containing 10 mM triethanolamine, 0.15 M NaCl pH 7.8, 0.5% NP-40, 10 mM iodoacetamide, 1 mM phenylmethylsulphonyl fluoride, 1 mM EDTA and 1  $\mu$ g ml<sup>-1</sup> leupeptin. Rabbit anti- $\gamma$  chain serum<sup>7</sup> (15  $\mu$ l) was added with 25  $\mu$ l of protein A coupled to Sepharose 4B (Pharmacia) and the mixture was incubated for 2 h at 4 °C with mixing. The precipitates were washed six times with the above buffer and analysed by reducing SDS-PAGE on a 12.5% acrylamide gel. Lane 1 and 2, OE4, lanes 3 and 4 CD8<sup>+</sup> IEL. Lanes 1 and 3, anti-CD3, lanes 2 and 4, anti- $\gamma$  chain serum. Lane 5,  $M_r$  markers shown  $\times 10^{-3}$ .

These results suggest that at least a portion of IEL were activated *in vivo* by as yet unknown antigens.

IEL may represent several novel T lymphocyte lineages but a precise function for IEL *in vivo* and in disease states has yet to be demonstrated. A large percentage of CD3<sup>+</sup>8<sup>+</sup> IEL do not express Thy-1 with the percentage of Thy1<sup>+</sup> cells ranging from 25 to 50% (data now shown). Moreover, a majority of these cells use  $\gamma$ ,  $\delta$  TCRs. (This paper and unpublished data based on separation of Thy-1<sup>+</sup> and Thy-1<sup>-</sup> IEL.) It is an interesting possibility that this lineage of T cells may not be thymic in origin, although the loss of Thy-1 could occur post-thymically. The anatomically restricted distribution of CD8<sup>+</sup> T cells that use  $\gamma$ ,  $\delta$  TCRs suggests that the differentiation pathway for these cells is distinct from that of other peripheral CD8<sup>+</sup> T cells. This possibility is strengthened by the observation that IEL are the only peripheral T cells that express the CT carbohydrate antigens. The CD4<sup>+</sup>8<sup>-</sup> dendritic epidermal cells that carry  $\gamma$ ,  $\delta$  TCRs may be a third lineage of T cells, although it remains possible that these cells are somehow linked to IEL in their differentiation pathway. Considering the organ-specific distribution of T cells with  $\gamma$ ,  $\delta$  TCRs it is likely that the antigenic reactivity of these TCRs will be different from that of  $\alpha$ ,  $\beta$  TCRs. Unlike other peripheral CD8<sup>+</sup> T cells, freshly isolated IEL could be induced to display cytolytic activity by engaging the TCR with anti-CD3 mAb, indicating that activation of IEL had occurred *in vivo*. T cells that carry the  $\gamma$ ,  $\delta$  TCR can exhibit non-MHC restricted cytolytic activity<sup>5,28-31</sup> and MHC-linked, but not MHC-restricted, cytolytic activity<sup>32</sup>. The effector populations in all cases were CD4<sup>+</sup>8<sup>-</sup> and had been activated *in vitro*, however. Whether IEL recognize foreign antigen in the context of Class I MHC molecules, as would be consistent with CD8 expression, remains to be seen. Learning the specificity of CD8<sup>+</sup> IEL will be an important aspect of deciphering the significance of  $\gamma$ ,  $\delta$ -TCR expression *in vivo*. The ability to isolate large numbers of



**Fig. 4** Cytotoxic activity of CD8<sup>+</sup> IEL. CD8<sup>+</sup> IEL or lymph node cells were isolated by panning on anti-CD8 coated petri dishes. Serial dilutions of effector cells were incubated in 96-well round-bottomed microtitre plates with  $5 \times 10^3$  of the following <sup>51</sup>Cr-sodium chromate-labelled target cells: P815 (DBA/2 mastocytoma), EL4 (H-2<sup>b</sup> thymoma) or YAC-1 (H-2<sup>a</sup>, NK-sensitive). Anti-CD3 mAb was added to a final concentration of 1  $\mu$ g ml<sup>-1</sup> (where indicated). ▲, P815 plus anti-CD3; △, P815 without anti-CD3, CD8<sup>+</sup> IEL as effectors; ●, P815 plus anti-CD3, CD8<sup>+</sup> lymph node cells as effectors. No significant lysis of EL4 or YAC-1 cells was observed using CD8<sup>+</sup> IEL or lymph node cells at an effector:target ratio of 100:1. Assay time was 4 h. Percent specific lysis was calculated as  $100 \times [(c.p.m. \text{ released with effectors}) - (c.p.m. \text{ released alone})] / [(c.p.m. \text{ released by detergent}) - (c.p.m. \text{ released alone})]$ . Spontaneous release was <10% for all target cells.

IEL bearing  $\gamma$ ,  $\delta$  TCRs will allow a detailed examination of the ontogeny and function of this intriguing population of T lymphocytes.

We thank Dr John Coligan (NIAID) for the generous gift of the anti- $\gamma$  chain serum and Dr Jeffrey Bluestone (University of Chicago) for providing the anti-CD3 mAb and for helpful discussions. We also thank Dr Lynn Puddington and Ann Berger for reviewing the manuscript.

Received 30 March; accepted 18 May 1988.

- Hedrick, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 531-535 (1985).
- Kronenberg, M. *et al. Nature* **313**, 647-653 (1985).
- Marrack, P. & Kappler, J. *Adv. Immun.* **38**, 1-30 (1986).
- Brenner, M. B. *et al. Nature* **322**, 145-149 (1986).
- Bank, I. *et al. Nature* **322**, 179-181 (1986).
- Weiss, A., Newton, M. & Crommie, D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6998-7002 (1986).
- Lew, A. M. *et al. Science* **234**, 1401-1405 (1986).
- Pardoll, D. M. *et al. Nature* **326**, 79-81 (1987).
- Bluestone, J. A., Pardoll, D., Sharrow, S. O. & Fowlkes, B. J. *Nature* **326**, 82-84 (1987).
- Nakanishi, N., Maeda, K., Ito, K., Heller, M. & Tonegawa, S. *Nature* **326**, 770-773 (1987).
- Lanier, L. & Weiss, A. *Nature* **324**, 268-270 (1986).
- Lanier, L. *et al. J. exp. Med.* **165**, 1076-1094 (1987).
- Yoshikawa, Y., Reis, M. D. & Mak, T. W. *Nature* **324**, 482-485 (1986).
- Koning, F. *et al. Science* **236**, 834-837 (1987).
- Stingl, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 4586-4590 (1987).
- Lefrançois, L., *J. Immun.* **138**, 3375-3384 (1987).
- Petit, A. *et al. Eur. J. Immun.* **15**, 211-216 (1985).
- Parrot, D. *et al. Ann. NY Acad. Sci.* **409**, 307-320 (1983).
- Lefrançois, L. & Bevan, M. J. *Nature* **314**, 449-452 (1984).
- Leo, O., Foo, M., Sachs, D. H., Samelson, L. E. & Bluestone, J. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1374-1378 (1987).
- Staerz, U., Kanagawa, O. & Bevan, M. J. *Nature* **314**, 628-631 (1985).
- Behlke, M. *et al. J. exp. Med.* **165**, 257-262 (1987).
- Ernst, P. B. *et al. Eur. J. Immun.* **15**, 216-221 (1985).
- Klein, J. R. *J. exp. Med.* **164**, 309-314 (1986).
- Ernst, P. B., Clark, D. A., Rosenthal, K. L., Befus, A. D. & Bienenstock, J. *J. Immun.* **136**, 2121-2126 (1986).
- Klein, J. R. & Kagnoff, M. F. *J. exp. Med.* **160**, 1391-1396 (1984).
- Leo, O. *et al. J. Immun.* **137**, 3874-3880 (1986).



28. Brenner, M. B. *et al.* *Nature* **325**, 689-694 (1987).
29. Borst, J. *et al.* *Nature* **325**, 683-688 (1987).
30. Moingeon, P. *et al.* *Nature* **325**, 723-726 (1987).
31. Reinherz, E. *Nature* **325**, 660-663 (1987).
32. Matis, L. A., Cron, R. & Bluestone, J. A. *Nature* **330**, 262-264 (1987).
33. Lefrançois, L. *J. Immun.* **139**, 2220-2229 (1987).
34. Cerf-Bensussan, N., Quaroni, A., Kurnick, J. T. & Bhan, A. K. *J. Immun.* **132**, 2244-2252 (1984).
35. Markwell, M. A. K. & Fox, C. F. *Biochemistry* **17**, 4807-4817 (1978).
36. Oettgen, H. C., Petty, C. L., Maloy, W. L. & Terhorst, C. *Nature* **320**, 272-275 (1986).
37. Wysocki, L. J. & Sato, V. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2844-2848 (1978).

## Transcription of the dystrophin gene in human muscle and non-muscle tissues

Jamel Chelly, Jean-Claude Kaplan, Pascal Maire, Sophie Gautron & Axel Kahn

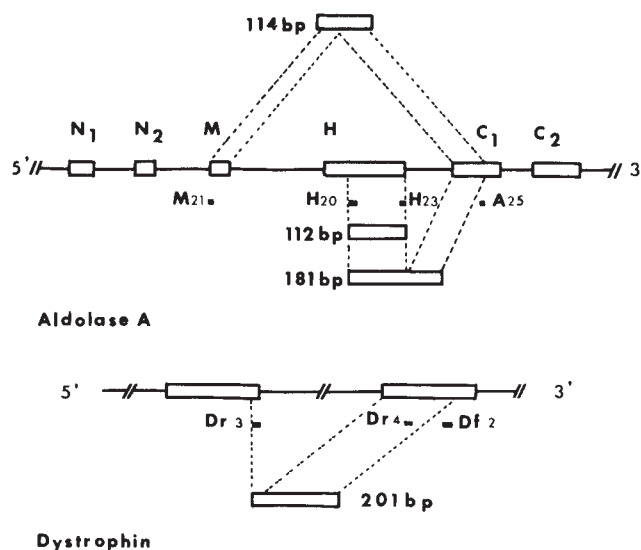
Unité de Génétique et Pathologie moléculaires, INSERM 129, and Laboratoire de Biochimie génétique, CHU Cochin, 24, Fg Saint Jacques, 75014 Paris, France

The gene that is defective in patients with Duchenne and Becker muscular dystrophy consists of about 60 short exons scattered along a gigantic DNA region that spans some 2 megabase pairs<sup>1,2</sup>. The encoded protein, dystrophin, was recently characterized as a component of muscle intracellular membranes of low abundance<sup>3,4</sup>. The dystrophin messenger RNA is difficult to study in both normal and pathological tissue specimens because it is large (14 kilobases) and scarce (0.01–0.001% of total muscle mRNA)<sup>2</sup>. We report here that efficient *in vitro* co-amplifications of the mRNAs of the dystrophin gene and of a reporter gene, aldolase A, by the polymerase chain reaction procedure<sup>5</sup> enables us to obtain a quantitative estimate of the dystrophin gene transcript. A processed, transcribed segment was thus detected in 13 different human tissues. It ranged from 0.02–0.12% of total mRNA in skeletal muscle to 25,000 times less in lymphoblastoid cells.

The first step was to synthesize a single stranded complementary DNA corresponding to a chosen segment of the dystrophin gene located in the DXS 206 region<sup>6</sup> which lies in the 5' part of the dystrophin coding sequence<sup>1</sup>. This was done by typical primer extension-reverse transcription in total RNA from skeletal muscle tissue using, as forward primer, a 24-mer oligonucleotide complementary to the 3' part of the mRNA segment to be amplified (see details in Fig. 2). In the next step, the second DNA strand was synthesized on the cDNA, and further amplification was obtained by the polymerase chain reaction (PCR) procedure<sup>5</sup>, using as reverse primer a 24-mer oligonucleotide of the same polarity as the mRNA sequence transcribed for the next exon upstream, 201 base pairs (bp) away from the forward primer<sup>1,7</sup> (see details in Fig. 1). To check the efficiency of the procedure, we simultaneously co-reverse transcribed and co-amplified another transcript as an internal standard in the same test tube. Aldolase A was selected because it has a tissue-specific transcription pattern, the two alternative exon H and exon M being used in housekeeping and in muscle-specific transcripts, respectively<sup>8</sup>.

Figure 2 shows that this procedure successfully amplified both dystrophin and aldolase A transcripts. The specificity of each amplified product was established: (1) by the size of the fragment stained by ethidium bromide; (2) by its specific hybridization in stringent conditions with the respective probes, sequence D<sub>r</sub>4 for dystrophin, and exon H, cloned in M13, for aldolase A (Fig. 1); (3) because no sequence was amplified when a forward dystrophin primer was used for reverse transcription, and when aldolase A primers located in a same exon (H23 and H20, see Figs 1 and 2b) was used for amplification.

The same procedure was used on total RNA isolated from



**Fig. 1** Positions of the different primers of aldolase A<sup>8</sup> and dystrophin<sup>1,7</sup> and size of respective cDNA fragments amplified. The sequence of the synthetic oligonucleotides was: A25: 5'CAGCTCCTTCTTCTGCTGCGGGGTC3', forward primer complementary to a segment of the first coding exon of aldolase A; H23: 5'CTGGCACAGGAGAGGGGCGGGTG3', forward primer complementary to a segment of non-coding exon H of aldolase A; H20: 5'CGCAGAAGGGGTCCTGGTGA3', reverse primer identical to a segment of non-coding exon H of aldolase A; M21: 5'ACTCGCTGTGACCAAGGCTCT3', reverse primer identical to a segment of non-coding exon M of aldolase A; D<sub>r</sub>2: 5'CTCCATCAATGAAGTCCAAATGA3', forward primer complementary to cDNA sequence 1,194–1,217 given by Koenig *et al.*<sup>1</sup>; D<sub>r</sub>3: 5'CATCAAATGCACTATTCTCAACAG3', reverse primer identical to cDNA sequence 1,017–1,040 given by Koenig *et al.*<sup>1</sup>, corresponding to cDNA sequence 415–438 given by Heilig *et al.*<sup>7</sup> (according to Heilig *et al.* they are different exons). D<sub>r</sub>4: 5'AAGAGCTATGCCTACACACAGGCTG3', a sequence between D<sub>r</sub>2 and D<sub>r</sub>3, used to probe the region amplified<sup>1</sup>.

nine other human tissues: fetal stomach, fetal intestine, fetal brain (cortex), fetal kidney, fetal and adult lung, fetal and adult liver, fetal and adult spleen, adult heart, and placenta. In addition, three different types of cultured cells of non-muscle origin were investigated: fibroblasts, lymphoblasts (eight different cell lines) and hepatoma cells (cell line Hep G2). Amplified transcribed sequences of aldolase A and of dystrophin were obtained in all these samples. In muscle tissues (skeletal heart and smooth muscle) the amplified segment was seen on ethidium bromide staining; in all the non-muscle tissues investigated, it was seen only after probe hybridization (Fig. 2a). In the lymphoblastoid cell lines we confirmed that the amplified segment exhibited the restriction fragment sizes anticipated from the cDNA sequence, indicating that it did derive from a true dystrophin mRNA.

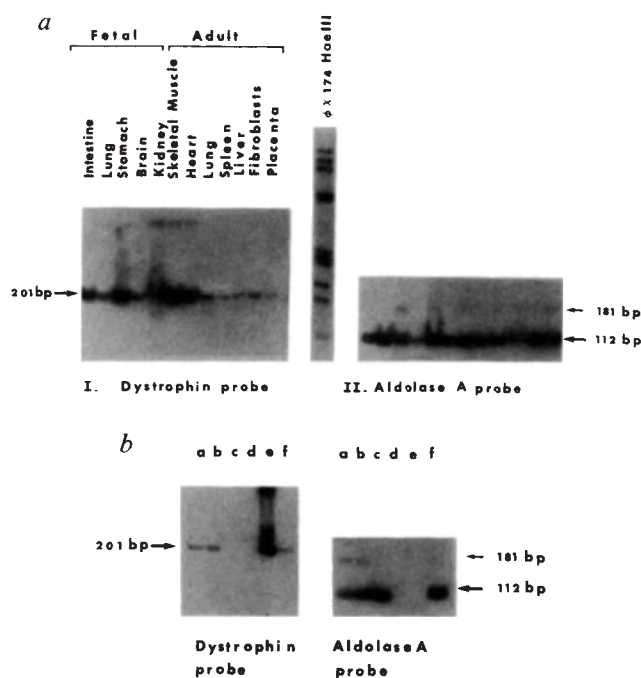
In contrast no dystrophin sequence was found in lymphoblastoid total RNA from a patient with DMD whose symptoms were glycerol kinase deficiency, adrenal congenital hypoplasia and mental retardation and who had a 4 Mb DNA deletion encompassing the whole DMD region<sup>9</sup> (Fig. 2b). The same negative result was obtained in lymphoblastoid total RNA from another DMD patient with smaller deletion of 250 kb encompassing pERT87.8 to XJ2.3 and spanning the amplified segment (data not shown). Aldolase A transcripts were normally amplified in RNA specimens from these patients, however (Fig. 2b).

The amount of dystrophin transcript obtained after amplification appeared to vary greatly from one tissue to another, and we tried to quantify our data. In the PCR method the amplification is exponential with a theoretical doubling of each strand at each cycle of amplification. In fact the efficiency

is less than 100%, and the extent of amplification ( $Y$ ) is given by the formula:  $Y = A(1+R)^n$ , where  $A$  is the initial amount of material,  $R$  the efficiency and  $n$  the number of cycles. It is possible to estimate  $Y$  with reasonable accuracy by measuring the extent of incorporation of  $5'$ - $^{32}\text{P}$  labelled primer into the amplified product, and  $R$  can be deduced from the slope of the semi-log plot  $\log Y = \log A + n \log (1+R)$ , as described in Fig. 3. For both co-amplified fragments (20–30 cycles), the efficiency remained constant in different runs. It was thus possible to calculate the ratio of dystrophin mRNA (or  $A_{\text{dys}}$ ) to aldolase A mRNA (or  $A_{\text{ald}}$ ) (Fig. 3). We found that  $A_{\text{dys}}$  was four times less abundant than  $A_{\text{ald}}$  in skeletal muscle, and eight times less abundant in smooth muscle (stomach). The amount of aldolase A mRNA has been estimated to represent  $\sim 0.1$ – $0.5\%$  of total mRNA in skeletal muscle, and at least ten times less in other tissues (S.G. *et al.* in preparation). From this value, we deduced the amount of dystrophin mRNA relative to total mRNA (Table 1).

Specific amplification of mRNA sequences, after a first step of reverse transcription, seems to be extremely sensitive. It has already been applied to total cDNA<sup>10</sup> and to a poly(A)<sup>+</sup> RNA fraction<sup>11</sup>. We detected a dystrophin-specific transcript in muscle tissues (skeletal muscle, myocardium, smooth muscle from stomach and intestine) and in non-muscle tissues (Table 1). By using two primers belonging to two different exons, we precluded any artefactual amplification of contaminating DNA (this was also excluded by the control experiment in which no amplification was observed when two aldolase A primers of the same exon were used without prior reverse transcription) and the signal with the expected size therefore necessarily corresponded to a dystrophin transcript that had been spliced, at least in this region. We found that within carefully checked experimental limits (exponential part of the curve, known yield of amplification) the starting material could be quantified. The amount of dystrophin transcript varied greatly, and a rough classification can be made as follows: group 1 (100%), skeletal muscle and heart; group 2 (5–10%), smooth muscle; group 3 (1–2%), brain cortex, kidney, lung; group 4 ( $\ll 1\%$ ), liver, placenta, spleen, fibroblasts and hepatoma cells. In lymphoblastoid cell lines the amount of dystrophin transcript was also in the group 4 range, but it fluctuated greatly from one cell line to another.

It has been reported that dystrophin mRNA is found only in skeletal muscle<sup>2</sup>, heart<sup>2</sup> and cultured myotubes<sup>12</sup>. These data were obtained after Northern blotting of poly(A)<sup>+</sup> fractions and



**Fig. 2** *a*, Autoradiograms of Southern blots performed on PCR co-amplified products of human transcripts of dystrophin and aldolase A genes hybridized with  $5'$ - $^{32}\text{P}$ -labelled Dr4 oligonucleotide (I) or H probe labelled by primer extension (II). The arrows indicate the size of the specific fragments amplified. With aldolase A the 181 bp fragment was concomitantly obtained because of the presence of residual A25 primer used for cDNA synthesis. *b*, Detection of aldolase A and dystrophin co-amplified transcripts in lymphoblastoid cell lines (LCL) and control experiments. Lane *a*, *b* and *f*: normal LCL; lane *c*: LCL from patient with 4 Mb deletion of the whole DMD region, showing no signal after hybridization with dystrophin probe, but the expected signal for aldolase A after hybridization with exon H probe; lane *d*: control of experiment showing no amplification from DNA (first strand cDNA synthesis of dystrophin, amplification with two aldolase A primers, H20 and H23, belonging to the same exon). Lane *e*: control experiment showing that in a sample treated as described in lane *d* amplification of dystrophin transcript is still obtained if two further dystrophin primers are added.

**Methods.** Co-amplification was as follows: (i) cDNA was synthesized by extension of forward primers ( $D_2$  for dystrophin, and A25 for aldolase A, see Fig. 1) in 5–10  $\mu\text{g}$  of total RNA<sup>18</sup> using reverse transcriptase (BRL) as described<sup>19</sup>; (ii) second strand synthesis and 20–30 cycles of co-amplification were using performed 2 u Taq-polymerase (Cetus) as described<sup>20</sup> after NaOH hydrolysis of RNA and EtOH precipitation. The primers used in this step (0.25  $\mu\text{g}$  each) were: reverse primers  $D_3$  for dystrophin and H20 for aldolase A, and forward primers  $D_2$  for dystrophin and H23 for aldolase A (Fig. 1). Temperatures were: 68 °C for polymerization, 90 °C for denaturation, and 25 °C for annealing. After electrophoresis and blotting the filter was hybridized, first with dystrophin probe, then, after stripping, with aldolase A probe. Exposure time was 3 h for aldolase A and 16 h for dystrophin. Full experimental details will be given elsewhere (J.C., in preparation).

**Table 1** Estimation of dystrophin transcript in different human tissues

| Tissue                                   | Fraction of dystrophin transcript: |                                   |
|------------------------------------------|------------------------------------|-----------------------------------|
|                                          | relative to skeletal muscle        | relative to total mRNA            |
| Skeletal muscle*                         | 100%                               | $2.5\text{--}12 \times 10^{-4}$   |
| Heart*                                   | 80–100%                            | $1.5\text{--}12 \times 10^{-4}$   |
| Smooth muscle:                           |                                    |                                   |
| stomach†                                 | 5–10%                              | $0.1\text{--}0.2 \times 10^{-4}$  |
| bowel†                                   | 2–5%                               | $0.04\text{--}0.1 \times 10^{-4}$ |
| Kidney†                                  | 1.6%                               | $0.03 \times 10^{-4}$             |
| Brain cortex†                            | 1%                                 | $0.02 \times 10^{-4}$             |
| Lung‡                                    | 1%                                 | $0.02 \times 10^{-4}$             |
| Liver (total and poly(A) <sup>+</sup> )‡ | 0.05%                              | $0.001 \times 10^{-4}$            |
| Placenta                                 | 1.05%                              | $0.001 \times 10^{-4}$            |
| Spleen‡                                  | 0.05%                              | $0.001 \times 10^{-4}$            |
| Cultured lymphoblastoid cells*           | 0.004–0.01%                        | 0.0001– $0.0002 \times 10^{-4}$   |
| Cultured fibroblasts*                    | <0.05%                             | $<0.001 \times 10^{-4}$           |
| Cultured hepatoma cell line* (Hep G2)    | <0.05%                             | $<0.001 \times 10^{-4}$           |

\* Adult origin; † fetal origin; ‡ both fetal and adult.

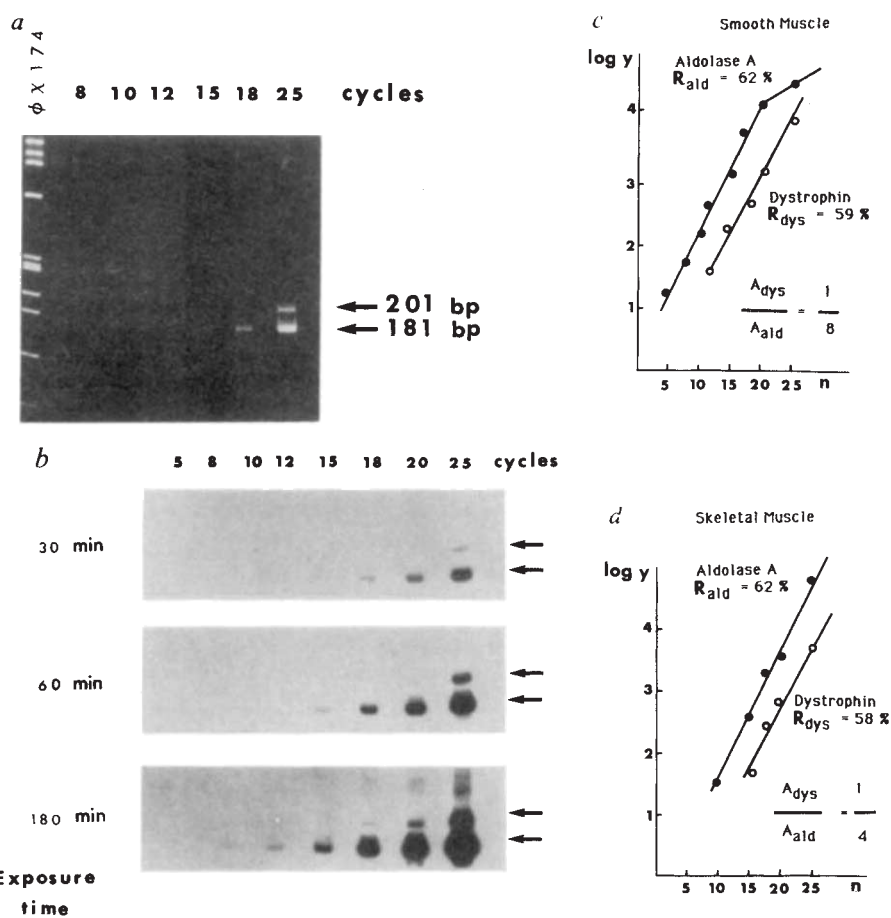
it is probable that this method greatly underestimates the actual amount of this huge mRNA. In fact it has been shown that mRNA species of large-size are preferentially lost in the oligo-dT procedure<sup>13</sup>. In contrast, the exquisite sensitivity of the PCR method, applied to total unfractionated RNA, enables us to find some dystrophin transcript in every tissue investigated. In particular its presence in smooth muscle and in brain is consistent with the recently reported finding of the dystrophin protein in both tissues, where the mRNA has not been detected by Northern blotting<sup>3</sup>.

The presence of extremely low amounts of dystrophin tran-



**Fig. 3** Quantification of co-amplified aldolase A and dystrophin transcripts in skeletal and smooth muscle. After the first step of cDNA synthesis from 10 µg of total RNA, each sample was divided into eight aliquots in which coamplification was performed with the 5' <sup>32</sup>P-labelled primers, H20 for aldolase A and D<sub>2</sub> for dystrophin (0.5 × 10<sup>6</sup> c.p.m. each). An increasing number of cycles was used as indicated. *a*, Ethidium bromide staining pattern of PCR co-amplified transcript fragments of aldolase A (181 bp segment between primers A25 and H20) and dystrophin (201 bp) from smooth muscle (stomach) after electrophoresis in 10% acrylamide. *b*, Similar gel to (*a*), dried and exposed for increasing times (30 min, 60 min, 180 min). *c*, Semi-logarithmic representation of the relative extent of amplification in smooth muscle, measured by counting the amount of <sup>32</sup>P incorporated into the fragments visualized in (*b*). *d*, Same plot as in (*c*) but from skeletal muscle (gel and autoradiography not shown).

**Methods.** The variable (*A*) is the amount of cDNA obtained in the first step, which depends on the amount of initial mRNA present in the sample. The absolute value of dystrophin mRNA (*A<sub>dys</sub>*) cannot be determined directly, but a relative figure can be deduced by comparing with the aldolase A internal standard (*A<sub>ald</sub>*) simultaneously reverse-transcribed and amplified in the same test tube. The relative value of *A<sub>dys</sub>* is calculated from the formula:  $A_{dys}/A_{ald} = Y_{dys} (1 + R_{ald})^n / Y_{ald} (1 + R_{dys})^n$  derived from the equations  $Y_{ald} = A_{ald} (1 + R_{ald})^n$  and  $Y_{dys} = A_{dys} (1 + R_{dys})^n$ . By measuring the amount of material amplified after each cycle, we found that the rate of amplification was exponential for 20 to 30 cycles, after which it decreased drastically and reached a plateau. These controls were done with aldolase A (112 bp in exon H), both from a cloned cDNA inserted in M13, and from total cellular LCL RNA (data not shown). The same kinetics was found after co-amplification of aldolase A and dystrophin transcripts in skeletal and smooth muscle. It was therefore possible to measure the slope  $\log(1 + R)$  of the plot  $\log Y = f(n)$ , and with the obtained value of the efficiency factor (*R*) to calculate the ratio *A<sub>dys</sub>*/*A<sub>ald</sub>*. In the other tissues dystrophin mRNA was amplified in the same conditions as in muscle, that is within the exponential part of the curve (20–25 cycles). The amplified product was hybridized with a <sup>32</sup>P-labelled oligonucleotide corresponding to a 25 nucleotide sequence internal to both primers. The intensity of the signal was estimated by densitometry and the value in different tissues compared to that obtained with skeletal muscle (Table 1).



script in some tissues such as liver, placenta or spleen is difficult to interpret. The possible contribution of smooth muscle cells from contaminating blood vessels is ruled out in cultured cells (fibroblasts, lymphoblastoid cell lines and hepatoma cell line Hep G2) (see Table 1). In these cells the dystrophin transcript represents about one copy per 500–1,000 cells, which cannot have any physiological significance. We think that this phenomenon may correspond to a general basal expression of any gene in any cell population. This is suggested by the fact that in lymphoblasts we also found a transcript of M-type aldolase A, although it is highly specific to adult fast twitch skeletal muscle fibres<sup>14</sup>. Transcripts of the highly liver-specific L-type pyruvate kinase gene<sup>15,16</sup> were also found in non-hepatic rat tissues, such as brain, lung, spleen (manuscript in preparation).

Due to its high sensitivity, the mRNA/PCR method is also suitable for addressing many other questions such as (1) possible alternative splicing patterns; and (2) the exploration of transcripts in muscle specimens from Duchenne patients. Indeed, preliminary results indicate that the dystrophin transcript is found in some patients (manuscript in preparation). Finally, it should be possible to explore normal and abnormal transcripts of any tissue-specific gene in any accessible cell.

After completion of this work a paper by Nudel *et al.* appeared<sup>17</sup>, in which, by a ribonuclease protection assay on total RNA, a dystrophin transcript was found in differentiated myogenic cell cultures and in brain.

We thank Dr Ketty Schwartz for providing an RNA preparation from normal adult heart, and Cetus for the generous gift of Taq polymerase. This work was supported by INSERM, Ministère de la Recherche et de l'Enseignement supérieur, and Association Française contre les Myopathies (AFM). J.C. is recipient of a fellowship from AFM.

Received 29 February; accepted 18 May 1988.

- Koenig, M. *et al. Cell* **50**, 509–517 (1987).
- Hoffman, E. P., Monaco, A. P., Feener, C. C. & Kunkel, L. M. *Science* **238**, 347–350 (1987).
- Hoffman, E. P., Brown, R. H. & Kunkel, L. M. *Cell* **51**, 919–928 (1987).
- Hoffman, E. P., Knudson, C. M., Campbell, K. P. & Kunkel, L. M. *Nature* **330**, 754–758 (1987).
- Saiki, R. K. *et al. Science* **230**, 1150–1154 (1985).
- Burghes, A. H. M., Logan, C., Hu, X., Belfall, B., Worton, R. G. & Ray, P. N. *Nature* **328**, 434–437 (1987).
- Heilig, R., Lemaire, C. & Mandel, J. L. *Nucleic Acids Res.* **15**, 9129–9142 (1987).
- Maire, P. *et al. J. molec. Biol.* **197**, 425–438 (1987).
- Chelly, J. *et al. Hum. Genet.* **78**, 222–227 (1988).
- Powell, L. M. *et al. Cell* **50**, 831–840 (1987).
- Veres, G., Gibbs, R. A., Scherer, S. E. & Caskey, C. T. *Science* **237**, 415–417 (1987).
- Lev, A. A., Feener, C. C., Kunkel, L. M. & Brown, Jr, R. H. *J. biol. Chem.* **262**, 15817–15820 (1987).
- McLeod, A. R., Karn, J. & Brenner, S. *Nature* **291**, 386–390 (1981).
- Keiichiro, J., Yuji, A., Tsunehiro, M. & Katsuji, H. *J. molec. Biol.* **190**, 401–410 (1986).
- Cognet, M., Lone, Y. C., Vaulont, S., Kahn, A. & Marie, J. *J. molec. Biol.* **196**, 11–25 (1987).
- Noguchi, T., Yamada, K., Inoue, H., Matsuda, T. & Tanaka, T. *J. biol. Chem.* **262**, 14366–14371 (1987).
- Nudel, U., Robzyk, K. & Yaffe, D. *Nature* **331**, 635–638 (1988).
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
- Schmid, A., Cattaneo, R. & Billeter, M. A. *Nucleic Acids Res.* **15**, 3987–3996 (1987).
- Saiki, R. K. *et al. Science* **239**, 487–491 (1988).

# Immunostaining of skeletal and cardiac muscle surface membrane with antibody against Duchenne muscular dystrophy peptide

Kiichi Arahata, Shoichi Ishiura, Tsuneo Ishiguro\*, Toshifumi Tsukahara, Yoshihiro Suhara, Chikahiko Eguchi\*, Tadayuki Ishihara†, Ikuya Nonaka, Eijiro Ozawa & Hideo Sugita

National Institute of Neuroscience, NCNP, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187, Japan

\* Central Research Laboratories, Ajinomoto Co., Inc., Kawasaki-ku, Kawasaki 210, Japan

† National Higashi-Saitama Hospital, Saitama 349-01, Japan

**Duchenne muscular dystrophy (DMD) is a debilitating X-linked muscle disease. We have used sequence information from complementary DNA clones, derived from the gene that is deleted in DMD patients<sup>1-3</sup>, to generate an antiserum that stains the surface membrane of intact human and mouse skeletal muscle, but not that of DMD patients and *mdx* mice<sup>4</sup>. Here we identify the protein reacting with this antiserum as a single component of relative molecular mass 210,000 ( $M_r = 210K$ ) that fractionates with a low-ionic strength extract of intact human and mouse skeletal muscle. It is therefore distinct from the 400K protein found in the heavy microsomal fraction of normal muscle and identified as a putative product of the DMD gene<sup>5</sup>. We also analyse further the**

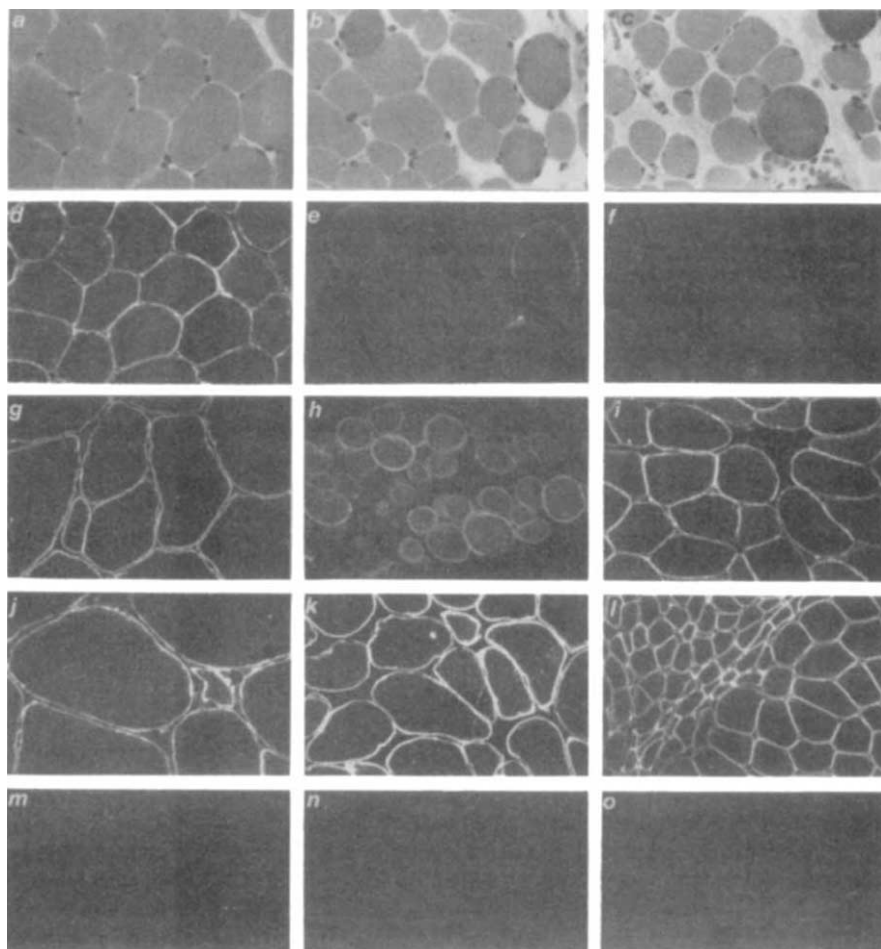
disease specificity of the antiserum. Positive staining is seen in normal controls, and in samples from patients with a wide range of muscular dystrophies other than DMD. Becker muscular dystrophy<sup>6</sup>, which is allelically related to DMD, was the only other exception, and gave a sporadic staining pattern. The demonstration of a specific defect in the surface membrane of DMD muscle fibres substantiates the hypothesis<sup>7-10</sup> that membrane lesions may initiate muscle degradation in DMD.

The sequence of the synthetic peptide (DMDP) used to raise the antiserum we describe is derived from the DMD complementary DNA (cDNA) clone, position 1,526 to 1,675. Our preliminary results with this antiserum suggested that the protein encoded by the DMD gene is localized at the surface membrane of normal muscle fibres and led us to the more extensive investigation we now report.

First, the reactions of skeletal muscle from different kinds of muscle diseases were compared with those of DMD (Table 1 and Fig. 1). Attention was focused on BMD which is thought to be the allelic disease to DMD<sup>6</sup>, although the relationship between the two diseases has been the subject of much controversy. Interestingly, the mode of staining of Becker muscular dystrophy (BMD) muscle membrane was different from that of either normal or DMD muscle (Fig. 1a-f) in that the surface membrane of BMD was only sporadically or faintly stained. The percentage of positive surface labelling seen in 1,000 randomly-selected muscle fibres in each of eight BMD cases ranged from 0.1 to 26.2%, and the mean of these eight cases (out of nine BMD patients) was 7.9% (s.d.  $\pm 10.4$ ). In one BMD case (that had been diagnosed by restriction fragment length polymorphism) positive labelling was observed in 91.3% of the

**Fig. 1** Frozen sections of biopsied skeletal muscle from various muscle wasting diseases stained with anti-DMDP antiserum using indirect immunofluorescence (d-l), together with sections stained by haematoxylin and eosin (H&E) (a-c), which correspond to samples used in d-f. Note the complete absence of immunostaining in DMD (c,f) compared with homogeneous staining of the membrane in all the normal muscle fibres (a,d). In patients with BMD, only partial staining of the surface membrane is observed (b,e). Emery-Dreifuss muscular dystrophy (g), congenital muscular dystrophy of Fukuyama type (h), limb-girdle muscular dystrophy (i), facioscapulohumeral muscular dystrophy (j), myotonic muscular dystrophy (k), and Kugelberg-Welander disease, a spinal muscular atrophy (l) show positive immunostaining of the surface membrane. ( $\times 130$ ).

**Methods.** Anti-DMDP antibody (IgG) was generated in a New Zealand white female rabbit as described previously<sup>4</sup>. Indirect immunofluorescence was used to stain cryostat sections on gelatinized coverslips<sup>15</sup>. Sections were preincubated with PBS containing 2% BSA and 5% heat-inactivated goat serum (solution G) then incubated with anti-DMDP antiserum ( $8 \mu\text{g ml}^{-1}$  in solution G) for 2 h at 37°C. After further incubation with the affinity-purified FITC-labelled goat F(ab')<sub>2</sub> anti-rabbit IgG (Tago Inc.,  $10 \mu\text{g ml}^{-1}$  in PBS/BSA) for 45 min at room temperature, the sections were mounted in a glycerol-based medium<sup>15</sup>, and observed in a Zeiss Axiophot microscope with epifluorescence. Specificity of the immunostaining was examined by replacement of the primary antibody with pre-immune IgG from the same rabbit (m), and pre-absorption of the primary antibody with DMDP (n). Both show negative immunostaining in normal muscle. In addition, complete blocking of the immunoreaction with unlabelled second layer antibody was observed (o). ( $\times 130$ )





muscle fibres. This suggests a partial or low-level and variable expression of the protein that is absent in DMD on the surface membrane of BMD muscle, and might explain the relatively benign and variable clinical course in comparison with DMD.

It is generally believed that other types of muscular dystrophy are not etiologically related to DMD. Indeed, all of the muscle specimens derived from the patients with 'dystrophies' other than DMD and BMD showed distinct reactions with the antibody in the surface membrane (Fig. 1g-l). Examples include Emery-Dreifuss muscular dystrophy, of which responsible gene locus is located on the long arm of X-chromosome<sup>11</sup> (Fig. 1g); congenital muscular dystrophy of Fukuyama type (Fig. 1h) and limb-girdle muscular dystrophy (Fig. 1i), both of which are transmitted as an autosomal recessive trait; and facio-scapulothoracic muscular dystrophy (Fig. 1j) and myotonic muscular dystrophy (Fig. 1k), which have autosomal dominant inheritance. Neurogenic muscular atrophy also showed the positive staining characteristic of normal muscle (Fig. 1l) as did all other myopathies so far examined (nemaline myopathy, myotubular myopathy, central core disease, inflammatory myopathies and hypothyroid myopathy, see Table 1). The fact that regenerating muscle fibres did not express the immunoreactive protein on the surface (Fig. 1h) may support the idea of developmentally regulated expression of DMD gene in myogenic cells<sup>12</sup>.

Next, tissues other than skeletal muscle were examined for immunoreactivity with the antiserum raised against DMDP (Fig. 2). As shown in Fig. 2, in accord with the results from skeletal muscle, the surface membranes of cardiac muscle from both normal human and control mice were immunostained with this antiserum, whereas those from DMD patients and *mdx* mice did not react with the antibody at all. Visceral smooth muscles from control mice showed a faint reaction, but other tissues so far tested, such as liver, kidney and brain, did not exhibit even a trace of reaction.

The 'membrane hypothesis', that is, the idea that some defect must exist in plasma membrane in muscle fibres of DMD, is

**Table 1** Immunohistochemical analysis of DMDP expression in the surface membrane of human and mouse skeletal and cardiac muscles

| Clinical diagnosis<br>mouse strain              | No. of<br>cases | DMDP Expression    |                   |
|-------------------------------------------------|-----------------|--------------------|-------------------|
|                                                 |                 | skeletal<br>muscle | cardiac<br>muscle |
| Duchenne dystrophy                              | 22              | —                  | —                 |
| Pre-clinical DMD                                | 5               | —                  | ND*               |
| Female DMD (X; 5 trans-<br>location)            | 1               | —                  | ND                |
| Becker dystrophy                                | 9               | -/+                | ND                |
| Emery-Dreifuss dystrophy                        | 1               | +                  | ND                |
| Congenital dystrophy<br>(Fukuyama)              | 5               | +                  | ND                |
| Limb-girdle dystrophy                           | 5               | +                  | ND                |
| FSH dystrophy                                   | 8               | +                  | ND                |
| Myotonic dystrophy                              | 5               | +                  | ND                |
| Dermatomyositis (child)                         | 6               | +                  | ND                |
| Polymyositis                                    | 5               | +                  | ND                |
| Spinal muscular atrophy                         | 6               | +                  | ND                |
| Other diseases†                                 | 7               | +                  | ND                |
| Normal human muscle‡                            | 9               | +                  | +                 |
| <i>mdx</i> mice (C57BL/10<br>ScSn- <i>mdx</i> ) | 5               | —                  | —                 |
| Control mice (C57BL/10 ScSn)                    | 5               | +                  | +                 |
| Total                                           | 104             |                    |                   |

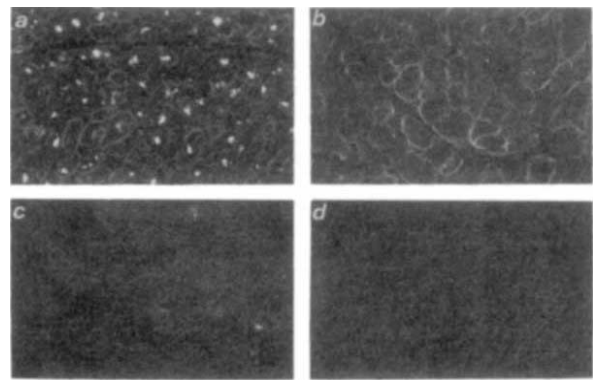
Tissues were prepared and examined as described in Fig. 1.

\* Not determined.

—, negative; +, positive -/+, immunostaining of the surface membrane was largely negative, but some muscle fibres show limited or diffuse membrane staining of the fibre as shown in Fig. 1.

† These include nemaline myopathy, central core disease, scleroderma and hypothyroid myopathy.

‡ Obtained from the autopsy cases which were free of muscle disease.

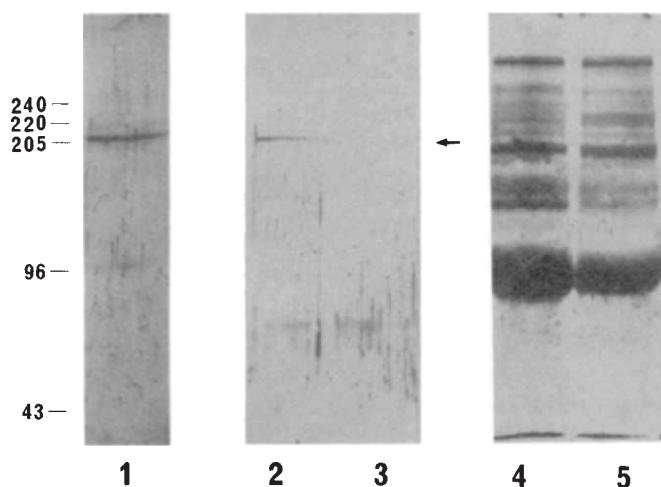


**Fig. 2** Human (a, c) and mice (b, d) frozen sections of cardiac muscles stained with anti-DMDP antibody by indirect immunofluorescence. Note the clear staining of all the surface membrane of both normal human (a) and control mouse (b) cardiac muscles. In marked contrast with this, there is no immunostaining of the surface membrane of both DMD (c) and *mdx* (d) cardiac muscle. Coarse granular spots observed in human cardiac muscles are yellowish autofluorescence materials. (×130)

**Methods.** Human cardiac muscles were obtained at the time of autopsy. Both *mdx* (C57BL/10 ScSn-*mdx*) and control B10 mice (C57BL/10 ScSn) were killed by cervical dislocation. Indirect immunofluorescence staining was as described in Fig. 1.

relatively old<sup>7-9</sup>. Such a defect is often present in nonnecrotic fibres, making small gaps in the plasma membrane in which overlying basal lamina is preserved<sup>7</sup>. A membrane defect would facilitate the leakage of cytoplasmic components, could lead to the high creatine kinase activities observed in the sera of patients with DMD<sup>13</sup>. It would also allow excessive influx of calcium into the fibre and may initiate the activation of calcium-activated protease in the muscle to cause the degradation of the fibre<sup>10</sup>. Until today, however, no specific molecular abnormalities of the muscle fibre surface membrane have been found. Our observations locate the deleted DMD protein at the surface membrane and have substantiated the membrane hypothesis at a molecular level.

Recently, Hoffman and his associates identified the protein product of DMD gene using polyclonal antibodies directed against peptide encoded by fragments of the mouse muscular dystrophy cDNA (refs 5, 14). According to them, this protein, named dystrophin, has a relative molecular mass of 400K and is included in the heavy sarcoplasmic reticulum fraction prepared by differential centrifugation. In our experiments, no material of *M<sub>r</sub>* 400K that cross reacted with anti-DMDP antiserum was found in crude cytosolic, myofibrillar and heavy microsomal fractions prepared from both human and rodent (mouse and rat) skeletal muscles according to the method of Hoffman *et al.*<sup>5</sup>. Instead, an immunoreactive component of *M<sub>r</sub>* 210K was found in the crude low-ionic strength extract (Fig. 3). This component could be coprecipitated from the extract with the anti-DMDP antiserum and migrated as a 210K protein on analysis by SDS polyacrylamide gel electrophoresis (data not shown). The 210K protein does not appear to be the degradation product of 400K dystrophin, because it was associated with the low-ionic strength extract of the myofibrillar fraction from both fresh and frozen mouse and frozen human skeletal muscle, but did not associate with the heavy microsome fraction containing triads, the site to which 400K dystrophin was localized by Hoffman *et al.*<sup>5</sup>. This suggests that our 210K protein is distinct from the 400K dystrophin, and emphasizes the possibility that the protein recognized by our antiserum is located in close contact with cytoskeletal components, probably the membrane associated cytoskeletal structures. As our antibody exclusively stained normal membranes of the muscle fibre, and did not react with those of DMD, the immunoreactive 210K protein certainly shares antigenic determinants with the DMD gene product. The



**Fig. 3** Western blot analysis of DMDP. Lane 1, normal human skeletal muscle; lane 2, control mouse skeletal muscle; lane 3, *mdx* mouse skeletal muscle. Lanes 4 and 5 are corresponding gels, stained with Coomassie Brilliant Blue (lane 4, control mouse skeletal muscle; lane 5, *mdx* mouse skeletal muscle).  $M_r$  standards ( $\times 10^3$ :  $\alpha_2$ -spectrins, myosin heavy chain,  $\alpha$ -actinin, actin) are shown in the left. An immunoreactive component of a molecular weight of 210K is not present in *mdx* mouse skeletal muscle, as shown in lane 3 (arrow). A non-specific 60K band is observed in lanes 2 and 3.

**Methods.** Crude DMD protein from human and mouse skeletal muscles were prepared as follows. Skeletal muscle (10.0 g) was homogenized with five volumes of 5 mM PBS and soluble and microsomal fractions were removed by ultra-centrifugation. The residual myofibrils were further washed three times with the same buffer to remove soluble proteins. After removal of myosin by Hasselbach-Schneider solution, the residue was kept in low-ionic strength solution for 8 h at 4 °C and centrifuged. The supernatant was subjected to ammonium sulphate fractionation. A 10–22 g per 100 ml ammonium sulphate precipitate of the extract (150  $\mu$ g) was subjected to 7% polyacrylamide gel in the presence of 0.1% SDS. After electrophoresis, the proteins were transferred to a nitrocellulose sheet and allowed to react with anti-DMDP antiserum. The reaction was visualized by ABC kit (Vector Laboratories, Burlingame). All the solutions for preparation contained 10  $\mu$ g ml<sup>-1</sup> leupeptin, 10  $\mu$ g ml<sup>-1</sup> E-64-c, 0.1 mM EDTA and 0.5 mM PMSF; SDS sample buffer also included 10  $\mu$ g ml<sup>-1</sup> phosphoramidon, 10  $\mu$ g ml<sup>-1</sup> elastatinal and 20  $\mu$ g ml<sup>-1</sup> bestatin to inhibit proteases.

lack of 210K protein in *mdx* mice is also consistent with the possibility that the 210K protein is a DMD gene product. The final identity of both 210K protein and 400K dystrophin with the DMD gene product, however, must await amino-acid sequence data.

In summary, we have identified a 210K protein, which is associated with the plasma membrane of normal muscle but absent from DMD patients and *mdx* mice. The specificity of our antiserum should provide an unparalleled tool for differential diagnosis of clinical and pre-clinical stages of DMD and other neuromuscular diseases.

We are very grateful to Professor S. Ebashi for his suggestions and encouragement and to Mmes K. Goto, E. Baba and Ms H. Anraku for their excellent technical assistance. Supported in part by grant-in-aid from the National Center of Neurology and Psychiatry (NCNP) of the Ministry of Health and Welfare, and by Scientific Research on Priority Areas, Ministry of Education, Science and Culture, Japan.

Received 6 May; accepted 19 May 1988.

1. Monaco, A. P. *et al.* *Nature* **323**, 646–650 (1986).
2. Koenig, M. *et al.* *Cell* **50**, 509–517 (1987).
3. Hoffman, E. P., Monaco, A. P., Feener, C. C. & Kunkel, L. M. *Science* **238**, 347–350 (1987).
4. Sugita, H. *et al.* *Proc. Japan Acad.* **64**, 210–212 (1988).

5. Hoffman, E. P., Knudson, C. M., Cambell, K. P. & Kunkel, L. M. *Nature* **330**, 754–758 (1987).
6. Kingston, H. M. *et al.* *Lancet* **ii**, 1200 (1983).
7. Mokri, B. & Engel, A. G. *Neurology (Minneapolis)* **25**, 1111–1120 (1975).
8. Schmalbruch, H. *Acta neuropath., Berlin* **33**, 129–141 (1975).
9. Carpenter, S. & Karpati, G. *Brain* **102**, 147–161 (1979).
10. Sugita, H. & Toyokura, Y. *Proc. Japan Acad.* **52**, 260–263 (1976).
11. Hodgson, S. *et al.* *Hum. Genet.* **74**, 409–416 (1986).
12. Nudel, U., Robzyk, K. & Yaffe, D. *Nature* **331**, 635–638 (1988).
13. Ebashi, S., Toyokura, Y. & Sugita, H. *J. Biochem., Tokyo* **46**, 103–104 (1959).
14. Hoffman, E. P., Brown, R. H., Jr. & Kunkel, L. M. *Cell* **51**, 919–928 (1987).
15. Arahata, K. & Engel, A. G. *Ann. Neurol.* **16**, 193–208 (1984).

## Immunoelectron microscopic localization of dystrophin in myofibres

Simon C. Watkins, Eric P. Hoffman\*, Henry S. Slayter & Louis M. Kunkel\*†

Structural Molecular Biology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts 02115, USA

\* Division of Genetics, Children's Hospital, and Harvard Medical School, Boston, Massachusetts 02115, USA

† Howard Hughes Medical Institute, Children's Hospital, Boston, Massachusetts 02115, USA

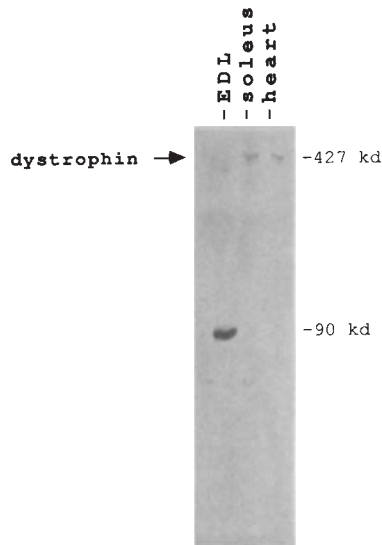
Duchenne muscular dystrophy, a common X-linked recessive human disease, has recently been shown to be caused by the deficiency of a large, low abundance protein called 'dystrophin'<sup>1–2</sup>. Biochemical techniques have shown dystrophin to be membrane-associated in skeletal muscle<sup>3</sup>, with enrichment of dystrophin in the t-tubules of 'triads'<sup>3,4</sup>. Other studies using immunohistochemistry on thick (10  $\mu$ m) sections have shown dystrophin to be located at the periphery of muscle fibres, possibly at the plasma membrane<sup>5–7</sup>. These results have been interpreted as being either consistent and complementary<sup>6</sup>, or contradictory<sup>7</sup>. To localize dystrophin more precisely relative to these membrane systems we have employed highly sensitive and spatially accurate immuno-gold electron microscopy of ultra-thin (70–100 nm) cryosections. The major distribution of dystrophin was on the cytoplasmic face of the plasma membrane of muscle fibres, and possibly on the contiguous t-tubule membranes. The presented data, taken together with recently accumulated information regarding the primary structure of dystrophin<sup>8</sup>, suggests that dystrophin is a component of the membrane cytoskeleton in myogenic cells. Thus, myofibre necrosis in patients affected with Duchenne muscular dystrophy is likely the result of plasma membrane instability.

Dystrophin has been shown to have sequence and/or structural similarity to other, more abundant, cytoskeletal proteins<sup>1,8–10</sup>. It is therefore important to have controls which ensure that only dystrophin is visualized, rather than related cross-reactive protein species. The *mdx* mouse<sup>11</sup>, an apparent null mutant for dystrophin<sup>1,12</sup>, serves as an excellent control for such cross-reactivity, as the immunological signal corresponding to dystrophin should be present only in normal mice, whereas the signals corresponding to cross-reactive protein species will be present in both normal and *mdx* mouse strains.

As shown in Fig. 1, the affinity-purified rabbit antisera raised against native dystrophin antigen appears monospecific for dystrophin on immunoblot analysis of normal mouse soleus and heart. In the extensor digitorum longus (EDL) muscle, however, this antibody preparation recognizes a high-abundance cross-reactive protein species of relative molecular mass 90,000 ( $M_r$  90K) in addition to dystrophin<sup>1</sup>. This protein species is localized within the myofibrillar Z-line by both fluorescent and immunoelectron microscopy (data not shown).

Immunolocalization of dystrophin in ultra-thin cryosections of both normal and *mdx* mouse soleus is shown in Fig. 2. Gold particles indicative of dystrophin immune complexes are seen along the inner face (cytoplasmic side) of the plasma membrane in normal mice (Fig. 2a,c). In *mdx* mouse muscle no labelling of the myofibre plasma membrane is seen, though non-specific labelling of an endothelial cell is evident (Fig. 2b). This indicates





**Fig. 1** Specificity of affinity-purified rabbit anti-mouse dystrophin antibodies. Immunoblot analysis of total protein extracts (50  $\mu$ g) from the normal mouse muscles indicated, using the antibody preparation used for the immunolocalization of dystrophin, shows that the antiserum is monospecific for dystrophin in soleus and heart, but that it cross-reacts with an additional protein species in the fast-twitch extensor digitorum longus (EDL) muscle.

**Methods.** Rabbit antiserum raised against a native fusion protein encoding a 30K portion of dystrophin was prepared as previously described<sup>1</sup>. Antibodies directed against the bacterial portion of the fusion protein antigen were absorbed from the serum by passage over a Biorad AffiGel-10 column to which partially-purified bacterial trpE protein had been coupled. The resulting crude serum was then passed over a similar column to which partially purified trpE-30K<sup>1</sup> antigen had been coupled. After extensive washing, antibodies specifically directed against the 30K dystrophin antigen were eluted with 0.2 M glycine (pH 2.3). The eluate was dialysed against phosphate-buffered saline (PBS), and quantitated with the Biorad protein assay. Solubilization and electrophoresis of total muscle protein on 3.5%–12.5% gradient denaturing polyacrylamide gels, and dystrophin detection, was as previously described<sup>1</sup>, using alkaline phosphatase-conjugated goat anti-rabbit IgG (Sigma). Other antibody preparations raised in sheep against SDS-denatured dystrophin antigens have been shown to be highly mono-specific for dystrophin in immunoblot analysis in all muscle types<sup>1,12</sup>. Upon immunoelectron microscopy, these affinity purified sheep antisera were seen to cross-react strongly with membrane components of the sarcoplasmic reticulum in the *mdx* mouse (data not shown) and were therefore not used in this study.

that the myofibrillar plasma membrane in this section was freely accessible to the antibodies. The dystrophin immuno-gold labelling did not show localization to membrane domains overlying any specific subsarcomeric region. We did, however, occasionally observe periodicity of the dystrophin labelling in both longitudinal (Fig. 2c) and cross-sections (Fig. 2a). When this periodicity was apparent it corresponded to  $\sim 1/10$ th of a sarcomere ( $\sim 125$  nm).

To determine if dystrophin could also be localized on the contiguous t-tubule system, immunolocalization was carried out on oblique sections of mouse EDL muscle (the t-system of 'fast' muscle such as EDL is approximately twice as abundant as that of 'slow' muscle such as soleus<sup>13</sup>). In these oblique sections, significant portions of the t-tubule system should be present between the inter-myofibrillar mitochondria and the M band of the sarcomere (A/I junction) (Fig. 2d,e). Some gold particles were present over this region in normal mice, but not in *mdx* mice, suggesting the presence of dystrophin at the t-tubule. Immunolabelling of the t-tubules, though, was only occasionally observed. The observed low density and frequency of t-tubule labelling might be expected given the density of labelling on

**Fig. 2** The immuno-localization of dystrophin in normal and *mdx* mouse skeletal muscle. Panels a, b, c. Ultrathin cryosections (cross-sections, a, b; longitudinal section, c) from either normal (a, c) or *mdx* (b) mouse soleus skeletal muscle. Two adjacent myofibres (MF, Panels a, c; z = z-line; M = M-line) or a myofibre and adjacent endothelial cell (END, Panel b) are shown. Gold particles (5 nm) indicating the location of dystrophin immune complexes, are seen along the plasma membrane (open arrows) of the normal mouse myofibres (a, c), but not the *mdx* plasma membrane (b). The labelling is localized immediately below the plasma membrane and seems somewhat periodic in its distribution, the period being approximately equivalent to the predicted molecular length of dystrophin<sup>8</sup>. Panels d, e. Oblique sections of normal (d) or *mdx* (e) mouse extensor digitorum longus (EDL) skeletal muscle. The beginning of the A band is seen in the upper right hand corner of each panel; mitochondria are in the upper left hand corner. The expected positions of the sarcoplasmic reticulum (SR) and t-tubules (T) are indicated. Sparse immunogold labelling is observed over putative t-tubules in the normal mouse muscle. Magnification: 98,000 $\times$  (panel a); 120,000 $\times$  (panel b); 80,000 $\times$  (panel c); 60,000 $\times$  (panels d, e).

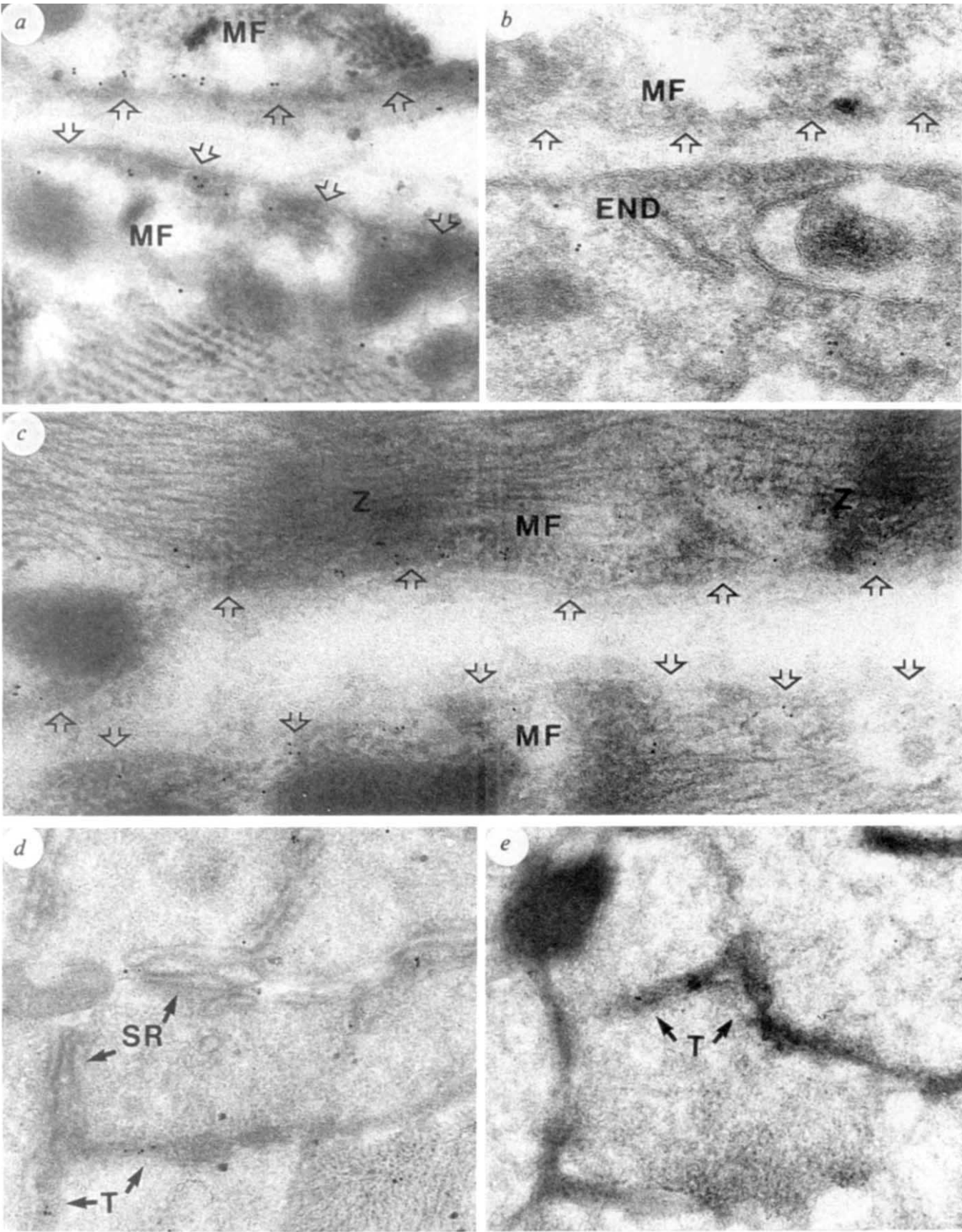
**Method.** The soleus and EDL muscles were carefully dissected from both normal and *mdx* mice, pinned to dental wax, lightly fixed (0.001%) glutaraldehyde, 2% formaldehyde) for 1 h at 4  $^{\circ}$ C, then prepared and sectioned as previously described<sup>20</sup>. Immunolabelling was found to be quenched when stronger fixation was used. Non-specific immunoreactivity was blocked by incubating the sections for 30 min in normal goat serum (5% in PBS/BSA) followed by three washes in PBS/BSA. Sections were incubated with affinity purified rabbit anti-mouse dystrophin (20  $\mu$ g ml<sup>-1</sup> for 30 min at room temperature), washed three times as above, incubated with a 5-nm gold-conjugated goat anti-rabbit second antibody (Janssen), and then washed three times in PBS/BSA. The remainder of the immunolabelling procedure was as described<sup>20</sup>. Sections were examined using a Philips EM300 electron microscope.

the external plasma membrane (Fig. 2) and the relative surface areas of these membrane systems<sup>15</sup>.

We have shown that dystrophin is sequestered immediately beneath the plasma membrane within the muscle fibre using immunoelectron microscopy of ultra-thin cryosections. Though we show evidence suggesting a similar association with the t-tubule membrane system, we feel this localization to be more suggestive than definitive. Nevertheless, an association of dystrophin with the t-tubules is strongly suggested by subcellular fractionation experiments employing sensitive biochemical markers for specific membrane systems<sup>4</sup>.

Recent analysis of the complete amino-acid sequence of the dystrophin molecule suggests a strong structural homology with the central rod domain of spectrin<sup>8,10</sup>—a large cytoskeletal protein involved in maintaining membrane structural integrity<sup>14,15</sup>. This domain has been predicted to be  $\sim 125$  nm in dystrophin<sup>8</sup>. Though dystrophin shows more extensive sequence similarity to a non-muscle isoform of alpha-actinin<sup>8,9</sup>, the role of this protein in the cytoskeleton is still poorly understood. The immunolocalization of dystrophin presented here supports the hypothesis of a 'spectrin-like' role of dystrophin in muscle membrane stabilization: first, we have found dystrophin to be localized on the cytoplasmic face of the plasma membrane, a localization similar to that of spectrin<sup>14</sup> and second, we occasionally observed a periodicity of the membrane-labelling of dystrophin in both longitudinal and cross-sections, suggesting a network of dystrophin molecules beneath the plasma membrane.

There is an extensive literature documenting plasma membrane defects which could lead to segmental necrosis of myofibres in DMD<sup>16–18</sup>. The data presented here suggest that such defects could indeed represent the primary cellular lesion resulting in the myofibre degeneration characteristic of DMD-affected muscle. If dystrophin contributes to plasma membrane structural integrity, it might be expected that myofibres undergoing the greatest physical stress would be those most





deleteriously affected by dystrophin deficiency. This correlation might explain the muscle group dependence<sup>19</sup>, age dependence<sup>12,19</sup>, and species dependence<sup>1</sup> of dystrophin deficiency on muscle fibre necrosis.

We acknowledge the generous support of the Muscular Dystrophy Association (USA) (LMK), and the National Institutes of Health (HSS and LMK). EPH is the Harry Zimmerman Post-doctoral Fellow of the Muscular Dystrophy Association. LMK is an Associate Investigator of the Howard Hughes Medical Institute.

Received 9 May; accepted 19 May 1988

1. Hoffman, E. P., Brown, R. H. & Kunkel, L. M. *Cell* **51**, 919-928 (1987).
2. Hoffman, E. P. *et al.* *New Engl. J. Med.* **318**, 1363-1368 (1988).
3. Hoffman, E. P., Knudson, C. M., Campbell, K. P. & Kunkel, L. M. *Nature* **330**, 754-758 (1987).
4. Knudson, C. M., Hoffman, E. P., Kahl, S. D., Kunkel, L. M. & Campbell, K. P. *J. biol. Chem.* (in the press).

5. Sugita, H. *et al.* *Proc. Japan Acad.* **64**, 37-39.
6. Bonilla, E. *et al.* *Cell* (in the press).
7. Zubrzycka-Gaarn, E. E. *et al.* *Nature* **333**, 466-469 (1988).
8. Koenig, M., Monaco, A. P. & Kunkel, L. M. *Cell* **53**, 219-228 (1988).
9. Hammond, R. G. *Cell* **51**, 1 (1987).
10. Davison, M. D. & Critchley, D. R. *Cell* **52**, 159-160 (1988).
11. Bullfield, G., Siller, W. G., Wight, P. A. & Moore, K. J. *Proc. Natn. Acad. Sci. U.S.A.* **81**, 1189-92 (1984).
12. Hoffman, E. H., Hudecki, M. S., Rosenberg, P. A., Pollina, C. M. & Kunkel, L. M. *Neuron* (in the press).
13. Eisenberg, B. R. in *Handbook of Physiology, Section 10: Skeletal Muscle* (ed. Peachy, L. D.) 73-112 (American Physiological Society, Bethesda, 1983).
14. Lazarides, E. *Cell* **51**, 345-356 (1987).
15. Lazarides, E. & Capetanaki, Y. G. in *Molecular Biology of Muscle Development* (eds Emerson, C., Fischman, D., Nadal-Ginard, B. & Siddiqui, M. A. Q.) 749-772 (Alan R. Liss, New York, 1986).
16. Rowland, L. P. *Muscle Nerve* **3**, 3-20 (1980).
17. Lotz, B. P. & Engel, A. G. *Neurology* **37**, 1466-75 (1987).
18. Cullen, M. J. & Mastaglia, F. L. *Br. Med. Bull.* **36**, 145-152 (1980).
19. Engel, A. G. in *Myology: Basic and Clinical* (ed. Engle, A. G. & Banker, B. Q.) 1185-1240 (McGraw-Hill, New York, 1986).
20. Watkins, S. C., Samuel, J. L., Marotte, F., Bertier-Savalle, B. & Rappaport, L. *Circ. Res.* **60**, 327-336 (1987).

## An anti-sense chalcone synthase gene in transgenic plants inhibits flower pigmentation

Alexander R. van der Krol, Peter E. Lenting, Jetty Veenstra, Ingrid M. van der Meer, Ronald E. Koes, Anton G. M. Gerats, Joseph N. M. Mol & Antoine R. Stuitje

Department of Applied Genetics, Vrije Universiteit, de Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

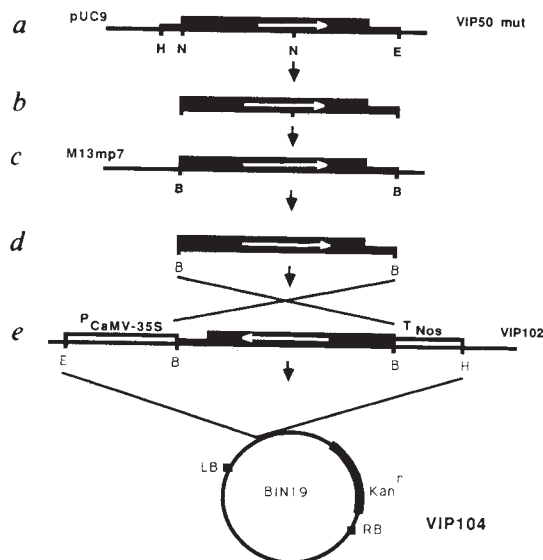
In most plants flower pigments derive from the flavonoid biosynthesis pathway. Consistent with this pathway in *Petunia hybrida* the key enzyme in flavonoid synthesis, chalcone synthase, is synthesized in the flower corolla, tube and anthers<sup>1</sup>. Here we show that constitutive expression of an 'anti-sense' chalcone synthase gene in transgenic petunia and tobacco plants results, with high frequency, in an altered flower pigmentation due to a reduction in levels of both the messenger RNA for the enzyme and the enzyme itself. The pattern of pigmentation varies among flowers of different transgenic plants, indicating that the activity of the anti-sense gene is influenced by DNA sequences that border its site of insertion in both a quantitative and a qualitative way. Backcrossing experiments show that the different pigmentation phenotypes resulting from the expression of anti-sense chalcone synthase gene(s) are stably inherited. These data establish that secondary metabolism in plants can be manipulated using transgenic plants that constitutively synthesize anti-sense RNA.

Recently strategies have been developed to mimic mutations by reducing gene expression using RNA that contains the complementary sequence to a given RNA, and hence is termed anti-sense RNA.<sup>2-7</sup> This approach has shown that the activity of introduced genes encoding nopaline synthase<sup>6</sup> and chloramphenicol acetyltransferase<sup>7</sup> can be effectively inhibited in plant and plant-cell model systems respectively. Here we describe the use of transgenic plants to study the effects of anti-sense RNA on the expression of endogenous genes, in this case those encoding chalcone synthase (CHS). Inactivation of expression of these genes can be scored phenotypically as a change in flower pigmentation, as CHS is the key enzyme in flavonoid biosynthesis. This system also enables us to monitor expression of the anti-sense gene in plant tissues that do not express the target gene(s).

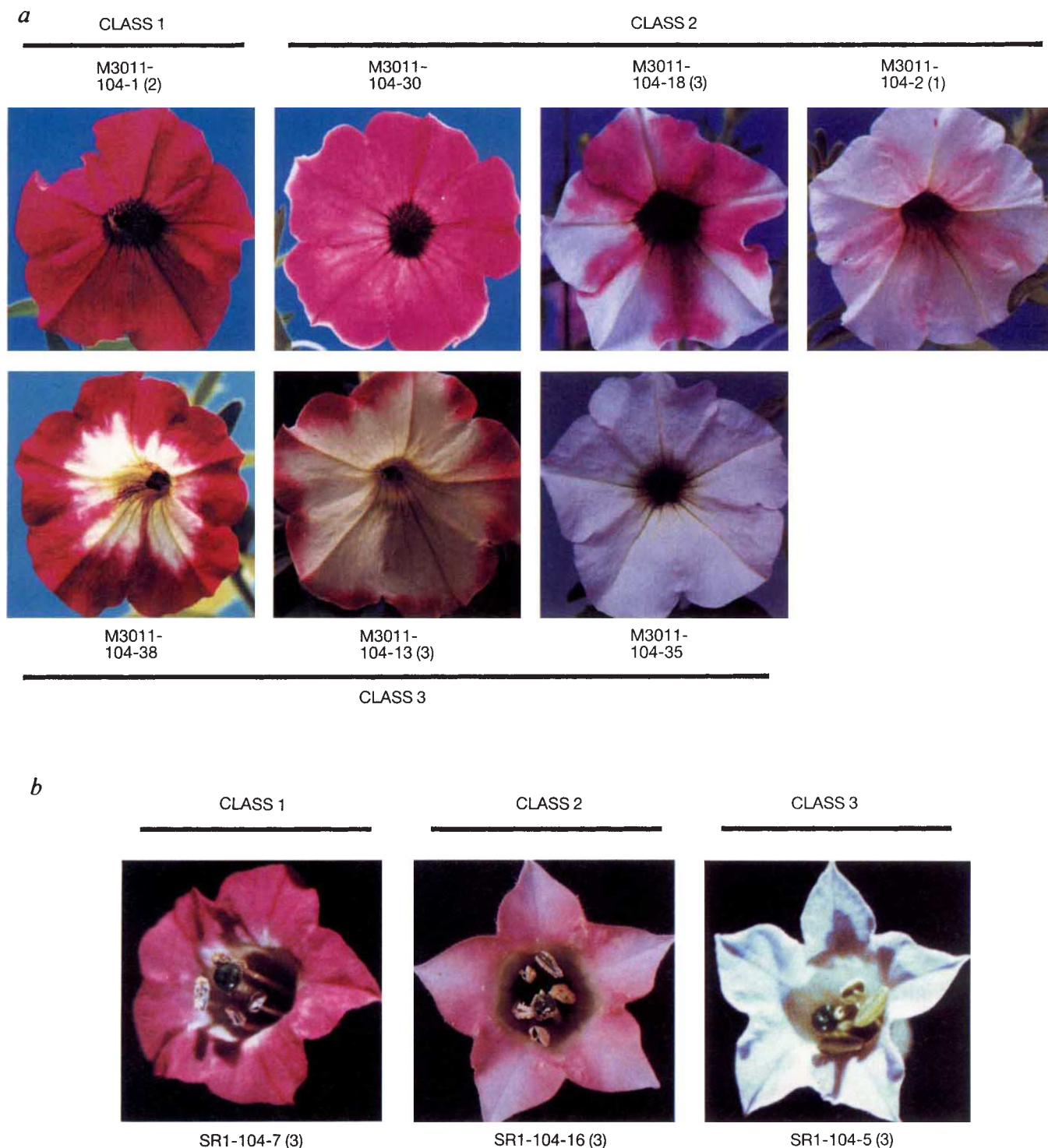
In *Petunia hybrida* CHS is encoded by a multi-gene family of which one member (gene A) is predominantly expressed in floral tissue<sup>1</sup>. There are nucleotide sequence differences between different petunia genotypes even within this particular gene<sup>1,8</sup>

and we chose the entire CHS gene A coding sequence from cultivar V30 as the template for construction of an anti-sense gene.

Previous experiments using anti-sense RNA suggest that a high ratio of anti-sense to sense mRNA is required for effective reduction of the expression of the target gene<sup>2,3,6,7,9</sup> and therefore we used the strong Cauliflower Mosaic Virus (CaMV) 35S-promoter<sup>10</sup> to direct the transcription of the anti-sense CHS gene in transgenic plants. Constitutive transcription from this promoter could thus create a pool of anti-sense RNA before the onset of endogenous CHS gene expression, providing an



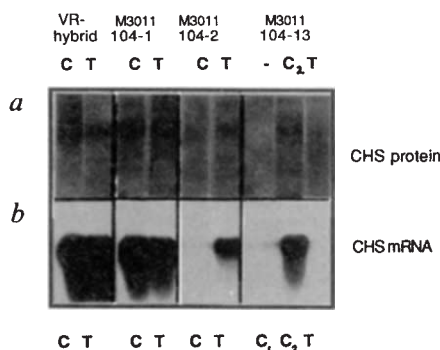
**Fig. 1** Construction of the anti-sense CHS gene (VIP104). A complementary DNA clone (VIP50 mut) of the petunia cultivar V30 CHS gene A with an introduced *Nco*I site at the initiation codon was used as the donor CHS coding sequence. *a*, VIP50 mut was partially digested with *Nco*I and completely with *Eco*RI; the sticky ends were made blunt using the Klenow fragment of DNA polymerase I and dNTPs. *b*, The resulting blunt-ended fragment was cloned into the *Hinc*II site of phage M13 mp7. *c*, The CHS coding sequence could now be isolated as *Bam*HI fragment. *d*, This fragment was cloned into the *Bam*HI site separating the CaMV 35S promoter and a nopaline synthase 3' flanking region that contains a poly(A) addition signal. *e*, A clone with the CHS coding sequence in reverse orientation behind the CaMV 35S-promoter (VIP102) was used to isolate the entire *Eco*RI-*Hind*III fragment and this was cloned into the polylinker site of the binary vector BIN19 (ref. 15) resulting in VIP104. *E*, *Eco*RI; *B*, *Bam*HI; *N*, *Nco*I; *H*, *Hind*III.



**Fig. 2** Flower pigmentation patterns of transformants of *Petunia* VR hybrid (**a**) and *Nicotiana tabacum* SR<sub>1</sub> plants (**b**). The flowers of transgenic petunia plants are divided into three classes. Those of tobacco transgenic plants also exhibit three phenotypes: 36 out of 40 had flowers indistinguishable from wild-type tobacco flowers (class 1); 3 plants gave flowers with sectorized pigmentation (class 2); and 1 plant gave entirely white flowers (class 3).

**Methods.** Plasmid VIP104 was mobilized into *Agrobacterium tumefaciens* strain 4404 using standard triparental mating techniques<sup>16</sup>. Plants were transformed using the leaf disc transformation method<sup>17</sup> and were regenerated on shoot-induction medium containing 250 µg ml<sup>-1</sup> kanamycin to select for cells expressing the neomycin phosphotransferase (NPT) gene contained in the T-DNA region of vector BIN19. Transformation of each plant was confirmed by Southern blot analysis of genomic DNA as follows. Total DNA was isolated from each transformant<sup>18</sup>, 5 µg DNA was digested with *Hind*III, electrophoretically separated on a 1% agarose gel and transferred to Genescreen Plus membrane (NEN Research Prod.). The blot was then probed with <sup>32</sup>P-labelled NPT DNA or <sup>32</sup>P-labelled CHS cDNA as described<sup>18</sup>. As digestion with *Hind*III should produce a unique fragment for each NPT copy integrated into the genome, the number of bands that hybridized with NPT was used as a measure of the copy number, given between brackets for some transformants.





**Fig. 3** Western blot (a) and Northern blot (b) analysis of class 1, 2 and 3 petunia transformants.

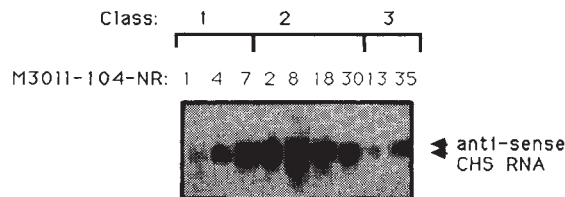
**Methods.** Between five and seven full-length but still closed flower-buds were harvested from each transformant. Tube and corolla tissue were separated and frozen in liquid nitrogen. Corolla tissue from transformant M3011/104/13 was separated into colourless ( $C_1$ ) and coloured ( $C_2$ ) parts. *a*, Western blots were made as described<sup>19</sup>. Equal amounts of protein extracted from either tube (T) or corolla (C) tissue, were used for each sample. CHS protein was immunologically detected using rabbit CHS anti-serum<sup>19</sup> and alkaline phosphatase conjugated to goat anti-rabbit IgG<sup>20</sup>. *b*, Northern blots on Genescreen Plus membrane from equal amounts (2  $\mu$ g) total RNA extracted from tube (T) or corolla (C) tissue were made as described (manuscript in preparation). CHS mRNA was detected using <sup>32</sup>P-labelled anti-sense CHS RNA transcribed by T7-polymerase from the pTZ18U vector (US Biochem. Corp.) containing the *Eco*RI–*Hind*III fragment of VIP50. Hybridization and washing of the blot occurred according to suppliers instructions, but at 75 °C instead of 60 °C.

instantaneous high ratio of anti-sense to sense CHS mRNA. Figure 1 shows the construction of the anti-sense CHS gene (VIP104) which we used to transform petunia VR hybrids<sup>11</sup> and *Nicotiana tabacum* SR<sub>1</sub> plants.

Normal petunia VR hybrid flowers have uniformly coloured corollas, tubes and anthers, but flowers of the petunia transgenic plants exhibited three different types of pigmentation patterns (Fig. 2a). Class 1 (12 out of 25 transformants) have flowers indistinguishable from wild-type VR hybrid flowers. Eight independent transformants in class 2 show a reduction of flower pigmentation in the corolla, often resulting in alternating coloured and colourless (white) sectors, although the tube remains coloured in these transformants. A similar pattern of expression has been reported in transgenic *Petunia hybrida* flowers that carry the A1 maize gene (which encodes another enzyme of the flavonoid biosynthesis pathway) also driven by the CaMV 35S-promoter<sup>12</sup>. In the five transformants in class 3 the colouration pattern seems to be fundamentally different from that of class 2 in that inhibition of flower pigmentation starts in the tube and extends outward into the corolla tissue, leading either to white flowers with a coloured ring or to entirely white flowers. Pigmentation of the anthers did not seem to be affected in any of the transgenic plants.

TLC analysis of extracts from white flower tissue showed an absence of flavonoids. The block in pigment biosynthesis could be bypassed, however, by feeding naringenin-chalcone, the product of CHS activity, to white corolla tissue. Thus the enzymes that operate after CHS in the pathway are present and active in these transgenic plants (results not shown).

Western (protein) and Northern (RNA) blot analyses using flowers of different classes showed that CHS protein levels and CHS mRNA steady-state levels in the tube and corolla correlate well with the observed phenotypes (Fig. 3a and b). Restriction analysis of the genomic DNA of some transformants did not reveal any changes in the endogenous CHS gene copies (data not shown). The altered pigmentation patterns described above were not observed in plants transformed with 'sense' CHS gene-



**Fig. 4** Northern analysis of anti-sense CHS-A gene expression in leaves of class 1, 2 and 3 petunia VIP104 transformants. The anti-sense CHS gene expresses two RNAs of about 1,300 and 1,100 base pairs (bp) in both petunia and tobacco transgenic plants. As a putative extra poly(A) addition signal (AATAAT) was found in the anti-sense orientation of the CHS sequence 220 bp upstream of the nopaline synthase poly(A) addition site, we assume the two RNAs result from polyadenylation at different sites.

**Methods.** Northern analysis was performed as described in the legend of Fig. 3. Anti-sense CHS RNA was detected using <sup>32</sup>P-labelled sense CHS RNA transcribed by T7 polymerase from the pTZ19U vector (US Biochem. Corp.) containing the *Eco*RI–*Hind*III fragment of VIP50.

constructs (20 transgenic petunia plants analysed). This makes it unlikely that these phenotypes are the result of somaclonal variation, homologous recombination or gene conversion events. The effectiveness of the anti-sense CHS gene-construct is demonstrated by the fact that CHS gene expression can be inhibited even in a heterologous plant system, that is, when tobacco plants are transformed with the petunia anti-sense CHS gene construct (Fig. 2b, analysis not shown).

A number of petunia transformants were backcrossed to both VR parents (V23 and R51). The progeny of the backcrossed transformant containing one insert showed 53/43 (anti-sense/wild-type) segregation ( $\chi^2 = 1.04$ ,  $p = 0.31$ ). Moreover, Southern blot analysis shows the complete co-segregation of the anti-sense phenotype (pigmentation in sectors) with the insert (10 anti-sense and 10 wild-type plants analysed, data not shown). Chromosomal localization studies are in progress and will be published elsewhere.

It is generally believed that, as in prokaryotes, anti-sense gene effects in eukaryotes are caused by the formation of duplex RNA structures. Preliminary experiments do indicate the presence of such RNA duplexes in the flower tissue of some transformants. Unlike in prokaryotes, the presence of an anti-sense 5' leader sequence is not a prerequisite<sup>13</sup> in our system (see Fig. 1), suggesting that inactivation of gene expression is not at the level of initiation of translation. The actual mechanism of inactivation of gene expression by the anti-sense gene seems to be rather complex. First, there is no correlation between the number of anti-sense CHS gene copies introduced and the steady-state level of anti-sense CHS RNA in leaf tissue (Fig. 4, for example, transformant 18 versus 13). This confirms and extends the general observation that bordering sequences influence gene expression in a quantitative way—the so-called position effect<sup>14</sup>. Second, several independent transgenic plants exhibiting the same steady-state level of anti-sense CHS RNA in leaf tissue, show a difference in flower pigmentation (Fig. 4, for example transformant 7 versus 18). Third, some independent transgenic plants with a very low steady-state level of anti-sense CHS RNA in leaf tissue, show a pronounced effect on flower pigmentation (Fig. 4, transformant 13).

The different classes of flower pigmentation found are very striking. Assuming stable integration of the anti-sense CHS gene and a normal transcription of endogenous CHS genes in each transformant, the difference between flowers of class 2 and 3 indicates that the integrated anti-sense CHS gene(s) are expressed in fundamentally different ways. This must also be due to the position effect and it shows that bordering sequences may also influence gene expression in a qualitative way, perhaps

affecting the timing or spatial distribution of gene expression. Extensive analysis of anti-sense and sense CHS gene expression during flower development is in progress to elucidate the actual mechanism of the anti-sense gene effect. The results presented here demonstrate the potency of anti-sense RNA in manipulating secondary metabolite pathways in both a quantitative and a qualitative manner.

We thank J. J. A. Büsse for his excellent care of the plants. J. Meijer and W. Bergenhenegouwen for photography, H. Bartelsson for typing the manuscript and G. Mahler for inspiration. This research is supported in part by a grant from the Netherlands Organisation for the Advancement of Pure Research (NWO).

Received 25 January; accepted 4 May 1988.

1. Koes, R. E. *et al.* *Nucl. Acids Res.* **4**, 5229–5239 (1986).
2. Rosenberg, U. B., Preiss, A., Seifert, E., Jäckle, H. & Knipple, D. C. *Nature* **313**, 703–706 (1985).
3. Cabrera, C. V., Alonso, M. C., Johnston, P., Phillips, R. G. & Lawrence, P. A. *Cell* **50**, 659–663 (1987).
4. Boulay, J. L., Dennefeld, C. & Alberga, A. *Nature* **330**, 395–398 (1987).
5. Knecht, D. A. & Loomis, W. F. *Science* **236**, 1081–1086 (1987).
6. Rothstein, S. J., DiMaio, J., Strand, M. & Rice, D. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8439–8443 (1987).
7. Ecker, J. R. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5372–5376 (1986).
8. Reif, H. J., Niesbach, U., Deumling, B. & Saedler, H. *Molec. gen. Genet.* **199**, 208–215 (1985).
9. Izant, J. G. & Weintraub, H. *Science* **229**, 345–352 (1985).
10. Bevan, M. W., Mason, S. E. & Golet, P. *EMBO J.* **4**, 1921–1926 (1985).
11. Wallroth, M., Gerats, A. G. M., Rogers, S. G., Fraley, R. T. & Horsch, R. B. *Molec. gen. Genet.* **202**, 6–15 (1986).
12. Meyer, P., Heidmann, I., Forkmann, G. & Saedler, H. *Nature* **330**, 677–678 (1987).
13. Green, J. P., Pines, O. & Innoue, M. A. *Rev. Biochem.* **55**, 569–597 (1986).
14. Jones, J. D. G., Dunsmuir, P. & Bedbrook, J. *EMBO J.* **4**, 2411–2418 (1985).
15. Bevan, M. *Nucl. Acids Res.* **12**, 8711–8721 (1984).
16. Ditta, G., Stanfield, S., Corbin, D. & Helinski, D. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7347–7351 (1980).
17. Horsch, R. B. *et al.* *Science* **227**, 1229–1231 (1985).
18. Koes, R. E., Spelt, C. E., Mol, J. N. M. & Gerats, A. G. M. *Plant molec. Biol.* **10**, 159–169 (1987).
19. Mol, J. N. M. *et al.* *Molec. gen. Genet.* **192**, 424–429 (1983).
20. Tunen, A. J. van & Mol, J. N. M. *Archs Biochem. Biophys.* **257**, 85–91 (1987).

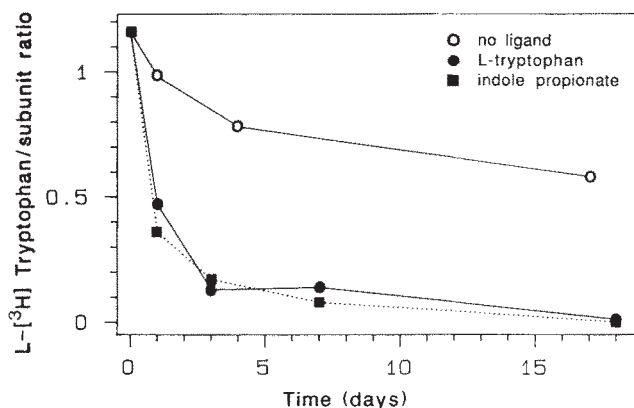
## The structure of *trp* pseudorepressor at 1.65 Å shows why indole propionate acts as a *trp* 'inducer'

Catherine L. Lawson\* & Paul B. Sigler†

Department of Biochemistry and Molecular Biology,  
The University of Chicago, Chicago, Illinois 60637, USA

The *trp* repressor is a small dimeric regulatory protein which controls the expression of three operons in *Escherichia coli*<sup>1–3</sup>. The inactive aporepressor protein must bind two molecules of L-tryptophan to form the active repressor<sup>4</sup>. If desamino analogues of L-tryptophan such as indole propionate (IPA) are substituted for L-tryptophan, an inactive pseudorepressor is formed<sup>1,5,6</sup>. Because the desamino analogues thus cause derepression of operons under control of the *trp* repressor, they appear to be 'inducers'. We have determined the crystal structure of the pseudorepressor and refined it to 1.65 Å. The molecular structure was compared to that of the nearly isomorphous orthorhombic form of the repressor. Surprisingly, the indole ring of IPA is in the same position as the indole ring of L-tryptophan in the repressor, but is 'flipped over'. As a result, the carboxyl group of IPA is oriented toward the DNA-binding surface of the protein and is in a position where it sterically and electrostatically repels the phosphate backbone of both operator and non-operator DNA. This explains why IPA acts as an apparent *trp* inducer.

The *trp* pseudorepressor is an inactive complex formed when



**Fig. 1** Ligand exchange in orthorhombic *trp* repressor crystals. Orthorhombic crystals of *trp* repressor were grown and stabilized as described<sup>12</sup>, but in the presence of L-[<sup>3</sup>H]tryptophan (Amersham). Selected well-formed crystals were transferred into 200 µl droplets of stabilizer containing either 1.75 mM L-tryptophan, 1.75 mM IPA or no ligand. The stabilizer was immediately removed and replaced twice to wash away excess radioactive ligand. Single crystals were removed at the time points indicated for analysis. The crystal was placed on a filter set on a sintered glass funnel and was quickly washed three times with 100 µl of stabilizer, with complete removal of each stabilizer wash by vacuum filtration. The filter was turned upside down into a 100 µl stabilizer droplet on a siliconized depression slide. The stabilizer was removed with a micropipette, and the crystal was dissolved in 100 µl of 100 mM sodium phosphate, pH 7.0. The protein content<sup>15</sup> and radioactivity of the dissolved crystal solution were measured.

the *trp* aporepressor binds indole analogues of L-tryptophan which, unlike L-tryptophan, confer no affinity for the *trp* operator. These indole analogues compete with L-tryptophan for binding *in vitro*<sup>7</sup>, and produce adducts that are similar in structure<sup>8,9</sup> but have lost their ability to bind operator. The effect of these analogues *in vivo* is presumably the same as they induce, or more precisely derepress, expression of the operons that are repressed by the *trp* repressor. The common features of these analogues are (1) an indole ring and a carboxyl group spaced with a variable length hydrocarbon chain and (2) lack of an  $\alpha$ -amino group. Interestingly, several desamino analogues bind with higher affinity to the aporepressor than does L-tryptophan<sup>7</sup>. These include indole 3-propionate (IPA), which is isostructural with L-tryptophan except for the replacement of the  $\alpha$ -amino group with a hydrogen atom.

Because desamino analogues of L-tryptophan do not act as corepressors, it has been postulated that the  $\alpha$ -amino group is critical for activation of the aporepressor protein<sup>1,5,6,10</sup>. From comparison of repressor and aporepressor structures<sup>11–12</sup>, it is clear that the binding of L-tryptophan is associated with a conformational change in the protein which allows two symmetry-related reading-head domains (D and E helices) to penetrate two successive major grooves of DNA. Because nearly isomorphous crystals of pseudorepressor (aporepressor plus IPA) can be obtained in at least two repressor crystal forms (ref. 8, and see below), the overall conformation of the active and inactive adducts must be quite similar. Also, nuclear magnetic resonance studies<sup>9</sup> suggest that there are no large conformational differences between repressor with L-tryptophan bound and pseudorepressor with IPA bound. As IPA produces the same global conformational change in the protein as L-tryptophan, the inability of IPA to confer operator affinity must be due to more subtle structural differences between the repressor and pseudorepressor.

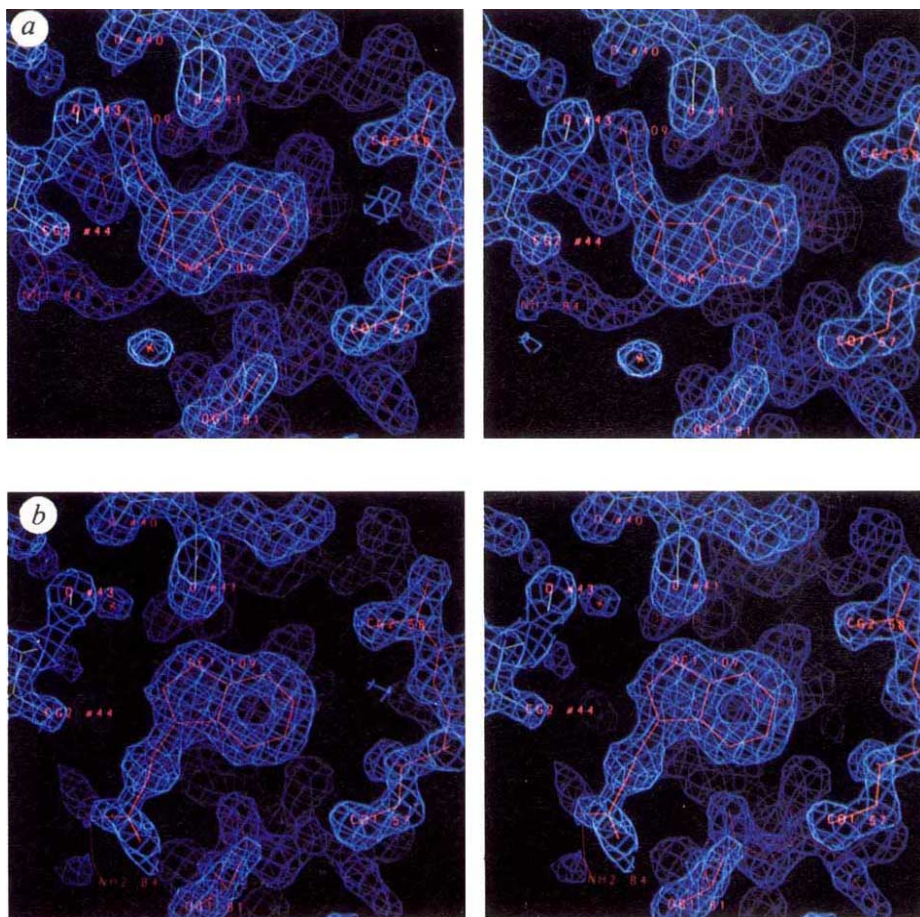
The orthorhombic crystal form of *trp* repressor will readily exchange the L-tryptophan ligand for IPA and other desamino analogues (Fig. 1). The ligand exchange is associated with modest intensity differences in the crystal diffraction pattern

\* Present address: Laboratory of Chemical Physics, Department of Chemistry, University of Groningen, Groningen, The Netherlands.

† To whom correspondence should be addressed.



**Fig. 2** Stereo views of the ligand binding site in the repressor and pseudorepressor structures generated using the program FRODO<sup>13</sup>. Both views are in approximately the same orientation. *a*,  $(2|F_{\text{obs}}| - |F_{\text{calc}}|)$ ,  $\alpha_{\text{calc}}$  map of the repressor, and *b*,  $(2F_{\text{obs}} - |F_{\text{calc}}|)$ ,  $\alpha_{\text{calc}}$  map of the pseudorepressor. Note that the indole group of IPA is 'flipped' with respect to the indole of tryptophan, and that the  $\alpha$ -amino group of the tryptophan is replaced by a water in the pseudorepressor structure. The side-chain of Arg84 is more poorly represented in the electron density map of the pseudorepressor structure than in the repressor. A '#' before the label number denotes a residue in the symmetry related subunit of the protein dimer.



that increase slightly with resolution. The structure determination and refinement of this crystal form of *trp* repressor to 1.65 Å resolution (crystallographic  $R = 18.0\%$ ) has been described elsewhere<sup>12</sup>.

Using phases from the refined orthorhombic *trp* repressor structure, a difference Fourier map with 'amplitudes' ( $|F_{\text{IPA}}| - |F_{\text{trp}}|$ ) from 9 to 1.65 Å was calculated and superimposed on the repressor model using the program FRODO<sup>13</sup>. The largest difference densities in the map ( $|\rho| \geq 5.0\sigma$ ) are clustered about the ligand. Strong positive and negative densities in the map indicated that the indole ring is bound in a similar way, but the aliphatic chain and carboxylate group of the IPA ligand in the pseudorepressor are positioned differently than those of the L-tryptophan ligand in the repressor. The initial model used for refinement against pseudorepressor data was generated by removing the L-tryptophan ligand and the side chain of Arg84 from the repressor model. This model was subjected to five cycles of least-squares restrained refinement using 5 to 2 Å data with the program PROFFT<sup>14</sup>. An  $(|F_{\text{obs}}| - |F_{\text{calc}}|)$  difference map was then calculated and the IPA ligand was modelled into the difference density. The positive density of the difference-map allowed an unambiguous interpretation of the conformation of the IPA ligand. The electron density of the aliphatic chain and indole group could only be satisfactorily modelled by positioning the indole group of the IPA such that it was 'flipped' with respect to the long axis of the indole group of the L-tryptophan ligand. The 'flip' leaves the six-membered ring in the indole of IPA in approximately the same position as that of L-tryptophan, consistent with the lack of ring density in the difference Fourier map. The pseudorepressor structure containing IPA modelled into the difference density was subjected to further refinement to 1.65 Å resolution. The data quality, coverage, restraints and crystallographic residual were comparable to those of the orthorhombic repressor ( $R = 18.2\%$ ). The striking difference in the

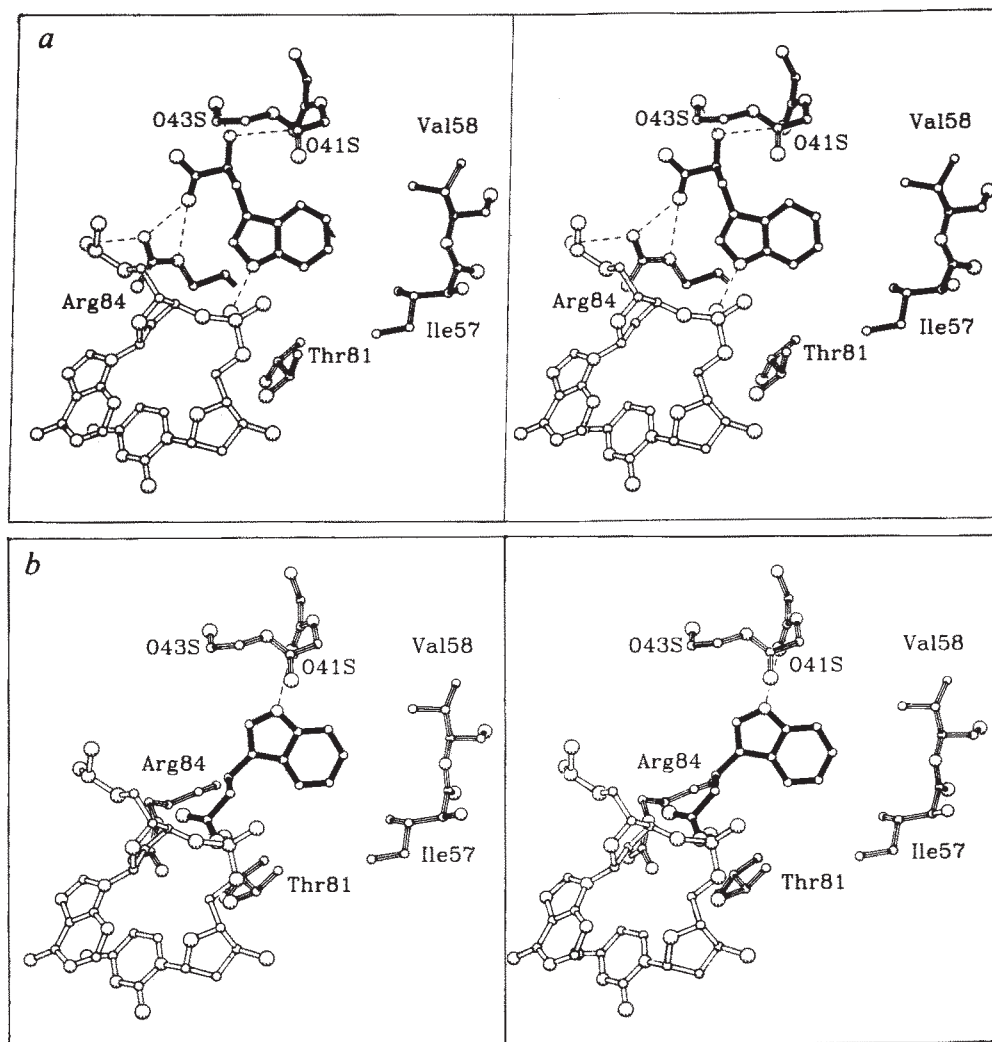
electron density of the ligation region of the repressor and pseudorepressor is shown in Fig. 2.

Figures 2 and 3 show the details of ligand binding. The  $\alpha$ -amino group of the tryptophan in the repressor structure forms hydrogen bonds with carbonyl oxygens of residues 41 and 43 in the carboxyl-terminus of the B helix, and presumably interacts with the  $-\frac{1}{2}$  charge of the helix dipole. The heterocyclic ring nitrogen of the tryptophan indole forms a hydrogen bond with an ordered water. In contrast, the heterocyclic ring nitrogen of the IPA of the pseudorepressor has been 'flipped' about a horizontal axis and forms a single hydrogen bond with the carbonyl oxygen of residue 41. An ordered water molecule of the pseudorepressor is positioned in the space occupied in the repressor by the  $\alpha$ -amino group of tryptophan and by hydrogen bonds to the carbonyl oxygens of residues 40 and 41. It is clear that the carboxyl-terminus of the B helix contributes to the compressor bending site a distribution of charge and potential hydrogen-bonds that plays a critical role in defining the stereochemistry of ligation.

The major effect of 'flipping' the ligand is to place the aliphatic chain of the IPA ligand in a completely different orientation than the corresponding chain in the L-tryptophan ligand (Figs 2b, 3b). The carboxyl group of IPA is not tucked back into the structure as in the case of L-tryptophan, but projects out towards the DNA binding surface. In both instances the guanidino group of Arg84 interacts with the carboxyl group of the ligand and is therefore positioned differently in the two structures. The rearrangement of the Arg84 side-chain is of questionable significance in the pseudorepressor because Arg84 forms a crystal contact which is not present in the repressor structure.

Marmorstein and coworkers<sup>7,10</sup>, in their systematic study of tryptophan analogues, conclude that binding a positively charged  $\alpha$ -amino group to the carboxyl terminus of the B helix is essential to support DNA binding for two reasons. First, the

**Fig. 3** DNA-contacts of the repressor and pseudorepressor in the ligand binding site. *a*, Interactions of the L-tryptophan corepressor that support DNA binding (Z. Otwinowski, *et al.*, in preparation). *b*, Interactions of the IPA 'inducer' that disrupt DNA binding. The repressor in the repressor/operator complex (Z. Otwinowski *et al.*, manuscript in preparation) was replaced with a model of the pseudorepressor by superimposition of corresponding  $\alpha$ -carbons. Note the direct clash of IPA's carboxyl group and a phosphate group of the operator.



amino group's positive charge neutralizes the negative pole of the helical dipole that would otherwise repel the nearby phosphate groups of the bound DNA. Second, by firmly anchoring the amino substituent and the indole ring in the indole binding pocket, the  $\alpha$ -carboxyl group can be placed in a sterically strained but critical orientation. The properly placed  $\alpha$ -carboxyl interacts through a charge-couple and two hydrogen bonds with the guanidino group of Arg84, which the recent crystal-structure determination of the *trp* repressor/operator complex shows is positioned to form a hydrogen bond with an operator phosphate (Z. Otwinowski *et al.*, manuscript in preparation and Fig. 3*a*). When the repressor structure is replaced (computationally) with a refined model of the pseudorepressor, it is clear that the guanidinium of Arg84 is no longer in a position to contact a backbone phosphate (Fig. 3*b*). What is most striking, however, is that the IPA is positioned in such a manner that its carboxyl group clashes with a neighbouring backbone phosphate. Thus, the absence of an amino group in IPA allows the 'inducer' to bind to the protein without stereochemical strain, but in a way that disrupts DNA binding in two ways. First, the presumed repulsion of DNA by the negative pole of the B-helical dipole is not neutralized. Second, and far more important, the carboxylate group no longer supports the interaction of the Arg84 side-chain with a backbone phosphate but, instead, causes both an electrostatic and steric clash with the DNA backbone. Because of this observation, we predict that the binding of desamino analogues of L-tryptophan to *trp* aporepressor affects binding in a non-specific manner; that is, the loss of operator affinity is due to a loss of affinity for DNA in general. Preliminary results with binding assays support this hypothesis (T. Wagner,

R. Q. Marmorstein, A. Joachimiak, L. Goldstein and P.B.S., manuscript in preparation).

We thank Professor Charles Yanofsky, Dr Richard Kelley and Ms Virginia Horn of Stanford University for the overproducing strains of *trp* repressor, Dr Barry Finzel of E. I. Du Pont De Nemours & Co. for his fast Fourier version of the Konnert-Hendrickson refinement program (PROFFT) and Professor Florante Quijoch and his colleagues at Rice University for PS300 FRODO 6.1. We thank Dr Richard Schevitz, Dr Andrzej Joachimiak, Mr Zdzislaw Otwinowski and Mr R. Q. Marmorstein for useful discussions. The research was supported by grants from the USPHS and the American Cancer Society. C.L.L. was a predoctoral trainee of the USPHS.

Received 11 April; accepted 19 May 1988.

- Rose, J. K., Squires, C. L., Yanofsky, C., Yang, H.-L. & Zubay, G. *Nature new Biol.* **245**, 133-137 (1973).
- Zurawski, G., Gunsalus, R. P., Brown, K. D. & Yanofsky, C. *J. molec. Biol.* **145**, 47-73 (1981).
- Kelley, R. L. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3120-3124 (1982).
- Joachimiak, A., Kelley, R. L., Gunsalus, R. P., Yanofsky, C. & Sigler, P. B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 668-672 (1983).
- McGeoch, D., McGeoch, J. & Morse, D. *Nature new Biol.* **245**, 137-140 (1973).
- Pauley, R. J., Fredericks, W. W. & Smith, O. H. *J. Bacteriol.* **136**, 219-226 (1978).
- Marmorstein, R. Q., Joachimiak, A., Sprinzl, M. & Sigler, P. B. *J. biol. Chem.* **262**, 4922-4927 (1987).
- Joachimiak, A., Schevitz, R. W., Kelley, R. L., Yanofsky, C. & Sigler, P. B. *J. biol. Chem.* **258**, 12641-12643 (1983).
- Lane, A. N. *Eur. J. Biochem.* **157**, 405-413 (1986).
- Marmorstein, R. Q. & Sigler, P. B. *J. biol. Chem.* (submitted).
- Zhang, R.-G. *et al. Nature* **327**, 591-597 (1987).
- Lawson, C. L. *et al. Proteins* **3**, 18-31 (1988).
- Jones, T. A. *Meth. Enzym.* **115**, 157-171 (1985).
- Finzel, B. C. *J. appl. Crystallogr.* **20**, 53-55 (1987).
- Bradford, M. M. *Analyt. Biochem.* **72**, 248-251 (1976).



# GUIDE TO AUTHORS

*Please follow these guidelines so that your manuscript may be handled expeditiously.*

*Nature* is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully.

Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

## CATEGORIES OF PAPER

**Review Articles** survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

**Articles** are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50–80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

**Letters to Nature** are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

**Commentary** articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

**News and Views** articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

**Scientific Correspondence** is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

**Submission.** All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

## GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

**Manuscripts** should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

**Titles** should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

**Artwork** should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

**Colour Artwork.** A charge of £500 a page is made for 4-colour figures. Inability to meet these costs will not prevent the publication of essential colour figures if the circumstances are explained.

**Figure legends** must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

**References** should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unrefereed abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963–65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

**Abbreviations,** symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

**Footnotes** should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

**Acknowledgements** should be brief. Grant numbers and contribution numbers are not allowed.